Une méthode pour l'évaluation de la qualité des images 3D stéréoscopiques.

Raluca Ioana Vlad

UNIVERSITÉ DE GRENOBLE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Signal, Image, Parole, Télécommunications (SIPT)**

Arrêté ministériel : 7 août 2006

Présentée par
Mlle Raluca VLAD

Thèse dirigée par
Mme Anne GUÉRIN-DUGUÉ et **Mme Patricia LADRET**

préparée au sein du
laboratoire Grenoble - Images, Parole, Signal, Automatique (GIPSA-lab)
dans **l'école doctorale d'Électronique, Électrotechnique, Automatique et Traitement du signal (EEATS)**

Une méthode pour l'évaluation de la qualité des images 3D stéréoscopiques

Thèse soutenue publiquement le **2 décembre 2013**,
devant le jury composé de :

Patrick LE CALLET
Professeur, Université de Nantes, Rapporteur

Frederic DUFAUX
Directeur de recherche, CNRS, Telecom Paris Tech, Rapporteur

Pascal MAMASSIAN
Directeur de recherche, CNRS, Université Paris Descartes, Examinateur, Président du jury

Ingrid HEYNDERICKX
Professeur, Technical University of Delft, Examinateur

Alain BELLON
Ingénieur, STMicroelectronics Grenoble, Examinateur

Anne GUÉRIN-DUGUÉ
Professeur, Université Joseph Fourier Grenoble, Directeur de thèse

Patricia LADRET
Maître de conférences, Université Joseph Fourier Grenoble, Encadrant de thèse

UNIVERSITY OF GRENOBLE

Doctoral school EEATS

(Électronique, Électrotechnique, Automatique et Traitement du Signal)

T H E S I S

for obtaining the title of

Doctor of Science

of the University of Grenoble

Mention: SIPT
(Signal, Image, Parole, Télécommunications)

Presented by

Ms Raluca VLAD

A Method for the Assessment of the Quality of Stereoscopic 3D Images

Thesis supervised by
Mrs Anne GUÉRIN-DUGUÉ and Mrs Patricia LADRET

prepared at
Grenoble - Images, Parole, Signal, Automatique Laboratory (GIPSA-lab)

presented on December 2, 2013

Jury:

Reviewers:	Patrick LE CALLET	-	University of Nantes
	Frederic DUFAUX	-	CNRS, Telecom Paris Tech
Examiners:	Pascal MAMASSIAN	-	CNRS, Paris Descartes University
	Ingrid HEYNDERICKX	-	Technical University of Delft
	Alain BELLON	-	STMicroelectronics Grenoble
Supervisor:	Anne GUÉRIN-DUGUÉ	-	University Joseph Fourier of Grenoble
Co-supervisor:	Patricia LADRET	-	University Joseph Fourier of Grenoble

Acknowledgments

The three years of my PhD studies have finished (it seems now, so fast) and the most important trace they leave with me is one of extreme human richness, coming from the wonderful persons I have been surrounded by. I am so lucky!

I am happy to address now my sincere thanks to Patricia and Anne, an excellent team of PhD supervisors, given their technicality, their precision, their high standards, but most of all, their admirable human qualities and their right words at the right time that have given me the confidence to overcome any challenge.

I respectfully thank my PhD committee members (Patrick Le Callet, Frédéric Dufaux, Ingrid Heynderickx, Pascal Mamassian, and Alain Bellon), for their kind appreciations of my work.

I particularly thank Alain Bellon of STMicroelectronics for his cordial support, for his very pertinent and detailed observations on my work and for the clear insight he allowed me to gain on the industrial side of my field of work.

I must thank Gelu Ionescu and Laurent Ott for their prompt technical help, but also for the cordial discussions that I have often had with each of them.

I thank Lucia Bouffard-Tocat and Isabelle Cieren for their precious help in administrative matters. They have made everything so much simpler for me.

I would like to thank Jérôme Mars, him being my first contact with GIPSA-lab and the one who believed in me when, back in Romania, I told him once how I would love to work in research related to image processing in France.

I owe many, many thanks to Simone and Fakhri. The numerous discussions with them on the subject of my PhD in our office have been so rich and have brought me so many good ideas.

I sincerely thank Olha and Matthieu for their contributions to the results presented in this manuscript. They have been both honestly involved in the experimental studies that we carried out together. Olha was very patient and calm when I was not anymore and Matthieu did an excellent job providing proper results in a very efficient way during my third year of PhD, allowing me thus to finish this manuscript correctly on time.

I must thank Damien for his proper and very elegantly organized LaTeX template that he allowed me to use, for all the LaTeX discussions (and jokes) he so often has to share, but also for the sincere and thorough help he has offered whenever I needed.

I have huge thanks to offer to my colleagues from the DIS department of GIPSA-lab: Arnaud, Aude, Céline, Christophe, Cindy, Emmanuelle, Francesca, Florian, Hao, Hélène, Quentin, Ladan, Jérémie, Jonas, Manu, Rémy, Robin, Rodrigo, Saman, Son, Vincent, Weiyuan (alphabetical order, impossible to do otherwise) and not only. They were all so supportive, so helpful, so kind, so sincere, that I consider them more than colleagues, I consider them my friends. I am so glad to have had the chance to work in such a pleasant environment.

I reserve particular thanks to Guanghan and Zhongyang, who have been my special com-

panions at GIPSA-lab since the beginning of my PhD and such devoted friends. I have shared with them so many great experiences.

I do not forget my friends from Romania either: Andra, Bobi, Cristina, Ioana, Monica. Even if we meet rarely and we sometimes do not get to speak even for months, whenever we do, it feels like we were never apart. I had their warm support when I decided to take the big step of coming to France and great encouragements in all the other steps that followed ever since. I can always count on them and I spend such pleasant moments in their company.

Cécile... Her friendship means so much to me. She is the first friend I made in France and definitely the most precious. I have learned so much from her, but most importantly I have learned that no effort is too big when you are doing it for a friend. Cécile, thank you!

Vincent... The proud smile on his face when I finished my PhD defense presentation, that was the beautiful final note of all the last months of emotion and hard work, during which he supported me like I did not know it was possible. Long discussions on so many subjects, insight from his experience, confidence, encouragement, sometimes tough (but justified and so helpful) remarks, typical (and cute) jokes, LaTeXor French orthography counseling at any hour, night or day, and so many other huge or very small things are some of the strongest reasons why I declare myself so lucky. Thank you, Vincent!

And, last, but most important, I would like to warmly thank my family: parents, grandparents, cousins, uncles and aunts. I am so thankful to have such a beautiful and loving family. They have supported me enormously in every step I took and what I am is because of them.

Thank you all !

Contents

Introduction

Introduction

As a direct consequence of the ever-growing interest in stereoscopic content, consistent research is being carried out at present, in order to better understand the perceptual mechanisms related to this type of content and in order to find methods for improving the quality of the production processes and of the final effect that stereoscopic systems have on viewers.

The problems encountered in the field are still numerous. To name only a few: the strong dependencies in the production chain that imply the necessity of a very elaborated control at every step in this chain, the viewing discomfort that the stereoscopic content may sometimes generate to the observer, or the lack of universal models or standards to allow the evaluation of the stereoscopic image quality and the comparison of stereoscopic systems or images among different studies. Our focus was on aspects related to this last point.

Thus, the present manuscript sums up the steps that we performed in our work, with the final goals of better understanding the judgment of the observers on the quality of stereoscopic content and of explaining this human perspective with algorithmic solutions. This objective is an extremely complex one and our work stands in the field as a step forward in the intricate labyrinth of understanding the human judgment on a multidimensional concept such as the stereoscopic image quality.

Our approach was part of a larger research project, called MOOV3D[1], which, as a partnership between the public research and the industry, has the objective of building a prototype mobile phone with a stereoscopic camera and with multiple alternatives for stereoscopic viewing: the local mobile screen, 3D glasses, or high-definition 3D television screens connected to the mobile phone. In this context, the stereoscopic quality needed to be studied in relation to the modeling of the human visual perception of depth, by including experimental investigations with human observers in simulated consumer utilization conditions, in order to guide the results towards the requirements of the project, strongly anchored in today's market reality.

Our focus was therefore on real photographed stereoscopic still images, for which the level of quality would be that rendered by existing consumer devices. The purpose was to spot the difficulties of acceptance in the state-of-the-art products and to propose quality models adapted to the present needs of the industry by identifying, validating, and quantifying the most important parameters that influence the stereoscopic quality. In our endeavor, we addressed the concept of quality in its absolute sense and not as relative to references.

The present manuscript will guide the reader through an introduction to the field of the stereoscopic image quality, then through a series of studies that we performed in order to shed more light on this subject.

Structure of the manuscript

The manuscript is structured in three different parts, all of them regrouping 11 chapters.

[1] MObile Original Video 3D

Part I is our state-of-the-art presentation of the context in which our work is placed.

In this first part, *Chapter 1* retraces the main points in the historical evolution of stereoscopy, in tight connection with the evolution of classical 2D images and motion pictures, from the first devices that could reproduce the natural sensation of depth to the systems of today.

Chapter 2 stands as an attempt of synthesizing the existing knowledge and insight that researchers have gained on the human mechanisms related to the perception of stereoscopic depth and on the connections between these mechanisms.

In *Chapter 3*, the different levels of the stereoscopic processing chain are explained, from capture, through compression, coding and data formats, up to display types, rendering, and visualization.

Ending Part I, *Chapter 4* brings into focus the premises of our work, by introducing notions on image quality in general and on stereoscopic image quality in particular. A review of the imperfections in the stereoscopic data that would influence quality is included, as well as observations on the added value of depth or on its negative effects, *i.e.* visual fatigue and visual discomfort. This chapter also tries to offer a global view on how the study of the stereoscopic quality can be approached. Several existing research results are mentioned, underlining in the same time the problems that remain open.

Part II introduces the objectives that we set in our work and the tools that we employed in order to reach these objectives.

The first chapter of this part, *Chapter 5*, is focused on an introductory presentation of these goals, on the steps to follow in order to accomplish them, and on our general methodological vision.

Chapter 6 gives detailed technical information on the stereoscopic system that we configured in our laboratory, to be used later for creating our own stereoscopic datasets.

And *Chapter 7* presents the experimental protocols that we implemented during our subjective experimental studies.

The totality of our results are detailed in *Part III*.

Chapter 8 presents the first subjective test that we implemented in the form of an exploratory study. The stereoscopic image quality model the we propose in this manuscript has its basis in the results of this particular study.

Chapter 9 shows our approach on refining the proposed stereoscopic image quality model by implementing new subjective tests, in the form of two quantitative experiments.

In *Chapter 10* we focused on our approach on adding the human factors perspective to our global view on the stereoscopic image quality, by performing a subjective test on the visual fatigue as a physical symptoms questionnaire.

The last of Part III, *Chapter 11*, concentrates all our previous experimental findings and all our algorithmic analyses. It details the steps we followed in order to propose a precise stereoscopic image quality model and in order to explain this model with objective measures.

A series of *Appendices* bring supplementary information, among which they give comprehensive details on the implementation of our experiments and they explain the mathematical or statistical tools that we employed in the processing of our results.

The results that we obtained, their significance, as well as their interpretation are summed up in our *Conclusions and Perspectives*, along with suggestions of future work directions that could bring even more precision to our results.

Part I

State-of-the-Art

Chapter 1

The Evolution of Stereoscopy

1.1 Short history of stereoscopy

The first major step towards the popularity that the 3D technologies enjoy today was made in 1833, when the first stereoscope was introduced by Sir Charles Wheatstone. The device he proposed used mirrors that helped the user fuse two slightly different pictures into one stereoscopic image, as illustrated in Figure 1.1. The observer was supposed to be placed in front of it, with the right eye in front of the mirror A and the left eye in front of the mirror A'. The pair of stereoscopic images fixed on the boards E and E' would thus be reflected in the mirrors and, when the D and D' panels were placed at the correct identical distances from the mirrors, each of the two images could be correctly seen by the corresponding eye and "the binocular image would be immediately seen single" [Wheatstone 1838]. The name chosen for the device was derived from the Greek *skopion* and *stereo*, meaning *solid sight* [Zone 2007], "to indicate its property of representing solid figures" [Wheatstone 1838].

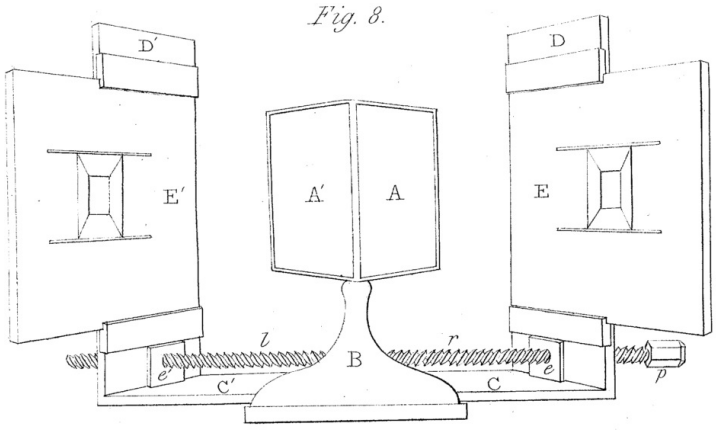


Figure 1.1: The stereoscope proposed by Sir Charles Wheatstone [Wheatstone 1838].

In 1838, Wheatstone further presented a paper called *Contributions to the Physiology of Vision, Part the First: On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision* [Wheatstone 1838] to the Royal Society of London, in which he detailed the theories he had elaborated on binocular vision. His exposition turned around the phrase: "the projection

of two obviously dissimilar pictures on the two retinae when a single object is viewed, while the optic axes converge, must therefore be regarded as a new fact in the theory of vision" .

The 1830*s* were revolutionary years for photography as well. The first *talbotypes* and *daguerrotypes* (inventions of William Fox Talbot, Nicéphore Niepce, and Louis Daguerre) allowed the fixation of photographic images on silver-chloride paper, respectively on silvered copper plates [Zone 2007]. The stereoscope started thus to use photographs, which replaced the initial paintings and drawings.

Sir David Brewster, a scientist well known for the invention of the kaleidoscope and a prolific writer, brought further contributions to stereography. He did extensive research on the historical bases of stereoscopy (going back to Euclid's *Treatise in Optics*), he redefined the rules of capturing stereoscopic images and was the first to suggest, in 1849, the idea of the twin-lens stereo camera [Lipton 1982]. But his most important contribution was in 1844, when he brought improvements to the Wheatstone stereoscope, transforming it into a smaller compact device that used magnifying lenses for viewing miniature stereoscopic cards, shown in Figure 1.2. The device had a horizontal opening S to allow the insertion of a slide containing two images, A and B, which, viewed through the tubes L and R, "instantly started into all the roundness and solidity of life" [Brewster 1856]. The lid $C - D$ allowed the admission of the necessary amount of light.

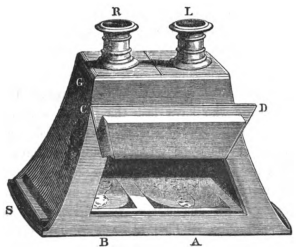


Figure 1.2: The lenticular stereoscope proposed by Sir David Brewster [Brewster 1856].

A second article by Wheatstone (*Part the Second*) [Wheatstone 1852], published in 1952, proposed as well a portable version of his original reflecting stereoscope and also a new type of device – the refractive stereoscope. A major part of the same article was dedicated to the utilization of the new photographic support with the stereoscope and to observations on the techniques to be used for the capture of stereoscopic photographs.

New improvements were brought to the stereoscope by Oliver Wendell Holmes, who made it easier to manipulate and cheaper to produce. It was him also who first used the term *stereography* (in Greek, *writing with solids*) in 1859 [Zone 2007]. Thus, soon the stereoscope reached considerable popularity both in Europe and in the United States, paving the path to what Ray Zone later called "the golden age of stereography from 1870 to 1920".

A contribution to this success of stereography was represented by the new devices that allowed the display of a sequence of stereo cards linked by a narrative. These *stereoview cabinets* have rapidly evolved between 1852 and 1875, especially in France, into elaborated "stereoscopic optical toys designed to achieve a synthesis of motion, color, and the third dimension" [Zone 2007]. Some of these were the *Stereorama*, the *Motoscope*, the *Bioscope*, the *Photobioscope*, or the *Stereoscopic Binocular Praxinoscope* [Lipton 1982]. As an example of such a stereoscopic moving picture *peep show*, the *Kinematoscope* proposed by Coleman Sellers, presented in Figure

1.3, had two openings on the sides that allowed the light to come in and, by watching through its two slits, a series of stereographs could be seen in succession while they turned around a central cylinder to which they were attached. All these devices had a strong impact on the public at first, but because of their technical imperfections and because of the dependency on a visualization device, they slowly disappeared from the general interest. They left their place to the various realizations of photography, which continuously improved at the time, without needing specific devices for viewing. However, all the *stereoscopic toys* played an enormous role in the evolution of stereography and are considered in the literature as the starting point for the later motion pictures [Zone 2007].

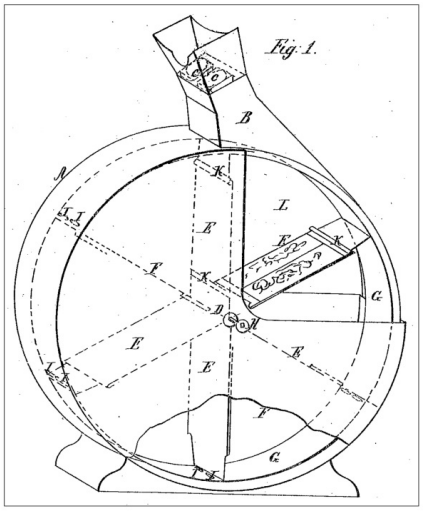


Figure 1.3: The Kinematoscope proposed by Coleman Sellers [Zone 2007].

At the end of the 1870*s*, Eadward Muybridge, hired by Leland Stanford at Palo Alto, stood out as one of the first researchers to combine stereoscopic techniques with what resembled the most to the motion pictures at that time – he is well known for his successive stereoscopic images of a galloping horse. Muybridge was followed by other pioneers, like Etienne-Jules Marey, with their efforts representing a proof of the importance of stereography for the study of motion [Zone 2007].

In 1889, William Kennedy-Laurie Dickson, one of Thomas Edison's assistants, developed the *kineto-phonograph*, a device that grouped together the projection of a motion picture on a screen with the sound of a phonograph. At the time, Edison showed less interest in the projection on a screen and focused more on patenting his *Kinetoscope* in 1893, under the name *Apparatus for Exhibiting Photographs of Moving Objects*. However, Dickson pursued his research into the motion picture projection and played an important role in the establishment of the 35-mm film format as a standard.

A significant motion picture pioneer was the British William Friese-Greene, who, in 1893, filed a patent for a system that recorded stereoscopic movies using a camera with two lenses and two negative films, the first camera to film "stereoscopic real-time photography of motion" [Lipton 1982]. The resulting movie could be seen through a specific stereoscopic device.

Regarding the means of separating the left and the right views during the projection of stereoscopic content, C.F. Jenkins was an important figure in the evolution of stereoscopy. In

1898, he filed a first patent describing a system of glasses with an electronic alternation of the left and right views, system that is still used at present. He was as well the founder of the *Society of Motion Picture Engineers (SMPE)* in 1919. Despite his fascination for stereoscopic images, in the 1920*s*, he declared his skepticism regarding their long-term success, given the practical difficulties of viewing them with various complicated devices [Zone 2007].

Wilhelm Rollman, in 1853, was the first to project stereoscopic images using color separation with red and blue lenses. He was inspired by the work of Heinrich Wilhelm Dove, who, in 1841, after developing a substractive printing process, managed to represent black and white images in relief, using the red and blue separation. Proposed by Louis Ducos du Hauron [Zone 2007], the word *anaglyph* (in Greek, *work in low relief*) was used to name this separation technique.

Polarization started to be used as a left-right stereoscopic separation method in 1891, as proposed by John Anderton, and both anaglyph and polarization have continued to be used until today. The polarization allowed the display of full-color images, however the anaglyph projection did not need the silver, non-depolarizing screen, specific to the polarization projections [Zone 2007], but it was an important cause of visual discomfort for a large number of viewers, and this still happens, even today.

The 1890*s* were critical years in the evolution of the motion pictures. The public projection started with the Skladanowsky brothers in Berlin (*Bioscope* films) and with the Lumière brothers in France (*Cinematographe* films) [Zone 2007]. In the meanwhile, *Arrivée d'un train*, the first stereoscopic motion picture, was presented to the public in 1935 by the Lumière brothers. The famous image of the train coming towards the public is shown here in Figure 1.4. Thus, the general interest in stereoscopic cinema appeared almost at the same time as cinema itself and its evolution followed closely that of classical cinematography [Devernay 2010].

Figure 1.4: Snapshot from the movie *Arrivée d'un train (à La Ciotat)* of the Lumière brothers [Lumière 1895].

The stereoscopic content projected at the time contained especially spectacular scenes, with strong visual effects, introduced to impress and attract the public. Unfortunately, for a long time, the realization primed over the content, keeping the stereoscopic cinema in a state of a *cinema of attractions* [Zone 2007]. The stereoscopic movies produced were however "sporadic - scattered all over the world and separated by long intervals of time" [Spottiswoode 1952]. The literature on the subject was also very limited, making the creation of good quality stereoscopic movies a real challenge.

In an attempt to compensate the lack of documentation, a text by Spottiswoode et al. was published in 1952 in the Journal of SMPTE [Spottiswoode 1952]. The article represents the

first important theoretical work on stereoscopy, giving ample technical recommendations and illustrating their utilization with examples from the filming of a stereoscopic movie produced that same year. However, his mathematical arguments were not taken into account on a large scale, being considered too complicated by some and judged sometimes wrong [Zone 2005].

In the meantime, in 1928, soon after the first stereoscopic movie in cinemas, the principle of stereoscopic television was demonstrated by John Logie Baird at the Baird Laboratories in London. Also, in the US, in 1953, the first experimental polarized 3D television broadcast took place and functioned correctly, but did not gain public success, since the homes were not equipped with the necessary compatible devices.

Following these disparate benchmarks related to the beginnings of the public interest in stereoscopy, two major *waves* of commercial success for stereoscopy in cinema were subsequently noted.

The first *wave* was in the 1950s, when a series of stereoscopic movies recalled the public attention to cinema halls, after a period of decline for the latter caused by the appearance of television. But, in the same time, the Cinemascope was also promoted and, from the two, the public preferred the second, while the interest in stereoscopic 3D was rapidly and almost completely lost. The main reason for this failure were the low standards of the stereoscopic movies of the time, both in terms of visual quality and content quality. Moreover, exaggerated effects were used in the movies, leading to increased visual discomfort. Nevertheless, about 65 stereoscopic movies were released in Hollywood between 1952 and 1954, among which *Bwana Devil*, the first feature-length commercial stereoscopic movie, was one of the top-grossing films of that time [Lipton 1982].

The second *wave* of success for the stereoscopic cinema, starting in the 1980s, brought, besides the standard stereoscopic movies, also movies in the IMAX 3D format. The first ones, of the same low quality as in the 1950s, had no more success than before. Nor had IMAX 3D however, despite the better quality of a few movies presented to the public in this format. The drawbacks that led to failure were mostly the technical difficulties and the very high costs implied.

In the same period, in 1982, experimental anaglyph 3D television broadcasts were made in European countries and numerous red/green stereoscopic glasses were sold for the occasion. Also, 3DTV demonstrations were done for the International Audio and Video Fair in Berlin in the 1980*s*, with equal considerable success. This showed that the public was interested in the 3D technologies and incited professionals from advertising, computer graphics, or meteorology to include stereoscopic solutions in their activities. The big limitation was however the lack of a convenient stereoscopic camera [Sand 1992].

1.2 Stereoscopy today

Since the second *stereoscopic boom* of the 1980*s*, researchers have turned towards stereoscopic 3D solutions in numerous and various fields. A review from 1992 gives examples of past stereoscopic projects related to neurosurgical operations, psychological research, endoscopy, or computer aided design (*e.g.* for televisions and theaters or in architecture) [Sand 1992]. The examples to this day are far more numerous than this, including television, smartphones, game consoles,

simulation systems, robotics, data visualization, medicine and biology, military applications, *etc.*

At present, we find ourselves in the middle of a new *stereoscopic renaissance* in cinematography, that began with the release of the animated movie *The Polar Express* in 2003, followed soon by *Chicken little*, *Beowulf*, and many others [Mendiburu 2009]. Thus, this time, it all started from animation, where the existing digital techniques allowed for a precise control of the stereoscopic parameters, producing content that ensured a comfortable and pleasant viewing. Live-action movies followed, among which *Avatar* was surely the most successful, promising good perspectives for the commercial future of 3D [Devernay 2010].

Hence, the appeal exerted by stereoscopy on the public proved in time to be certain. However, judging from the tumultuous history of stereoscopy, what is also certain is that in order for this technology to fully become integrated in our daily lives, we must still break several obstacles. The technical limitations (*e.g.* the unpractical stereoscopic cameras, the expensive infrastructure) must be overcome, the perceived stereoscopic 3D quality must be improved as much as possible by combining insights from algorithmic approaches and psycho-perceptual studies, and the visualization experience must be made pleasant and simple.

We think that the future is best synthesized by Mendiburu in his book *3D Movie Making – Stereoscopic Digital Cinema from Script to Screen*, appreciated as a reference in the movie-making community and probably the best technical introduction to the domain [Devernay 2010]. He says:

> Eventually, 3D will make its way into mainstream cinema the way color and sound did: it will be considered useless until it's available with a reasonable price tag. And then, all of a sudden, it will be unavoidable and ubiquitous, to the point that the very mention of "3D" will disappear from posters. At some point in the near future, you will go to see a "flattie" for nostalgia's sake, just as you sometimes watch black-and-white movies on TV today. [Mendiburu 2009]

This resolution will not materialize however until the stereoscopic content produced does not reach exceptional quality, which would impose mastering the mechanisms of human perception and judgment of stereoscopic images, but also the evaluation, the prediction, or the correction of their quality. These aspects will be discussed in the following chapters.

Chapter 2

The Human Stereoscopic Perception

Introduction

In the context of this work, which is centered on the *quality of stereoscopic images* such as perceived by *human observers*, it is essential to start from the *human* perspective, in the form of a detailed description of the functioning of the *human visual system (HVS)*. The present chapter has therefore the purpose of introducing the main notions related to the HVS and to explain its mechanisms considered as most important, with a special focus on the human perception of *depth*.

2.1 The main HVS characteristics related to the binocular vision

A series of physiological features of the HVS play important roles in the correct binocular perception. They are either directly related to the physical characteristics of the visual system, either to more complex brain processes triggered by various exterior stimuli, or to a combination of both. Sometimes, a strong emotional component is also involved in perception, linking objective data with our accumulated knowledge about the world.

2.1.1 Interocular distance

The main characteristic that makes the human binocular vision possible is the fact that the two eyes are placed at slightly different positions. This separation is measured between the two pupils and thus called *inter-pupillary distance* (IPD) or, less precisely, *interocular distance*. It varies from one person to another, being considerably shorter for small children, and it directly influences the way depth is perceived, in the sense that a small interocular distance generates larger perceived depths, opposed to a large interocular distance. The inter-pupillary distance values are usually comprised between 50 and 70 mm and the range can be extended to 40 to 80 mm in order to include children and extremes [Dodgson 2004]. The average value considered in most studies is however 63 mm or 65 mm.

2.1.2 Binocular disparity

In binocular vision, the separation between the left and the right eye allows for two distinct angles of view and thus for the capture of two images which represent slightly different perceptions of the world. These differences between the positions of the same points on the the two retinas are called *binocular disparities* and help create the sensation of depth when both retinal images are combined by neuronal processes.

The binocular disparity is formally defined as a difference of two angles and this is illustrated in Figure 2.1, where various aspects related to the geometry of the human binocular perception of depth are shown.

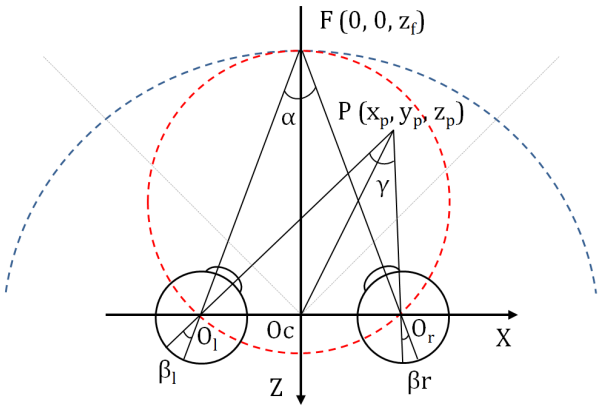


Figure 2.1: Geometrical aspects of the human binocular vision; adapted from [Liu 2008].

When fixating F, a given point in space, the axes of view of both eyes are oriented in the direction of that point, converging on it at a certain angle. Thus, that specific point has *zero disparity* and the points in front of or behind it in relation to the observer have *negative or positive disparities*, indicated by larger or smaller angles of convergence.

The binocular disparity is thus given by the difference between the angles that represent the deviations of the axes of each eye from the directions that give zero disparity. In the schema, the disparity of the point P, different from zero, can be expressed as:

$$d = \beta_r - \beta_l = \alpha - \gamma, \tag{2.1}$$

where

$$\alpha = 2 \cdot atan(\frac{-e/2}{z_f}), \tag{2.2a}$$

$$\varphi = atan(\frac{-x_p - e/2}{z_p}) - atan(\frac{-x_p + e/2}{z_p}), \tag{2.2b}$$

and where the e parameter represents the average interocular distance O_lO_r.

The *Vieth-Müller Circle* (VM circle) is illustrated by the red dashes in Figure 2.1 and represents the *theoretical horopter*, *i.e.* the totality of all the other points that can be perceived with zero disparity for a given fixation. This circle is the plane representation of the actual sphere which passes through the fixation point and the nodal points of both eyes [IJsselsteijn 2002b]. The points located inside the VM circle have *negative (crossed / near) disparity* and the points outside the circle have *positive (uncrossed / far) disparity.*

However, the real horopter does not coincide with the VM circle. The *physical horopter* represents the region of maximum stereoscopic precision given a certain fixation, while the VM circle is just the correct geometrical representation of the zero disparity points. The two are not necessarily identical in all situations, the real horopter generally having a much more complex shape than that of a circle [Blakemore 1970].

2.1.3 Vergence and accommodation

In order to visualize a stereoscopic stimulus, two main physical gestures are performed by the oculomotor system: the eyes converge and they set their focus on the stimulus. Therefore, the *vergence* and the *accommodation* are two mechanisms that play very important roles in the stereoscopic perception.

The *vergence* of the eyes is the rotation of their vertical axes, such that a certain point in space is fixated by matching the two retinal images in function of the existing disparities. This action is performed whenever the desired fixation point is located at non-corresponding points on the two retinas, or, otherwise formulated, whenever the disparity of the desired fixation point is different than zero. The purpose is to bring the fixated point in the center of both eyes' retinas [Watt 2013].

The *accommodative response* is the action of the ciliary muscles on the shape of the crystalline lens [Watt 2013]. It is performed when the eye changes fixation between points located at different depths in space, such that the new fixated point is in focus. Accommodation is triggered when, after a new fixation, the image on the retina is blurred (or, more precisely, of sub-optimal contrast) so it can be said that accommodation is *driven by blur* [Lambooij 2007].

Accommodation determines the retinal *depth of focus* (the acceptable range of focus on the

retina, measured as fractions of millimeters) and the corresponding *depth of field* (the acceptable range of focus in the visualized space in front of the eye, measured in meters). The two measures are often used interchangeably and abbreviated as *DOF* and they take equal values when expressed in vergences [Smith 1997]. However, for a given depth of focus, the interval of depth of field of the visualized scene varies in function of the distance to the fixated object: it is narrower for close objects and wider when viewing a scene from far away.

The required accuracy of accommodation is given by the eye's effective depth of focus, estimated to approximately $0.25 - 0.3\,\mathrm{D}$ [Watt 2013], where D represents the *dioptre* unit, equal to the reciprocal of the focal length in metres, thus a unit equivalent to 1/metres. In the meanwhile, the accuracy of vergence was estimated to an angle of approximately 15 to 30 minutes of arc [Watt 2013], corresponding to Panum's fusion area, discussed later, in Section 2.1.5.

In the stereoscopic perception of the real world, the two actions, vergence and accommodation, are correlated: when fixating an object in space, the accommodative response of the visual system automatically makes adjustments to bring that precise object into focus. This is called *vergence driven accommodation*, or simply *vergence accommodation*.

When watching stereoscopic 3D content on a screen, the HVS is challenged by the fact that the natural link between accommodation and vergence is broken: in order to perceive an object represented in depth, the eyes must focus on the display, while the fixation must be at a certain level of depth in front of or behind the display pane, where the object on screen is three-dimensionally represented. This is the case when *accommodative vergence* needs to occur. The eyes need to adjust their fixation for the specific depths represented on the display only after seeing the display in clear focus.

However, the natural limits of human vision allow to decouple the two ocular responses only to a certain degree and, consequently, a *zone of clear, single binocular vision* (ZCSBV) can be delimited [Watt 2013], beyond which the phenomenon of *accommodation-vergence rivalry* (or *accommodation-vergence conflict*) appears. This phenomenon occurs when the natural reaction of vergence accommodation leads to blurred or double images because of the conflict between the natural response of the oculomotor system and the requirements of a specific type of display or of a certain content displayed.

A much narrower limit can be considered as well, called *Perceival's zone of comfort* or just *zone of comfort* (ZoC), that represents the approximate range in which, despite the accommodation and vergence decoupling, the observed objects are correctly perceived, and, most importantly, without any discomfort [McCarthy 2010, Watt 2013].

Figure 2.2 and Figure 2.3 show the limits estimated in [Shibata 2011] for the ZCSBV and of the ZoC, the first in dioptres and the second in meters. In both images, the dashed diagonal line shows the real-world conditions of viewing, where accommodation and vergence are correlated and no rivalry phenomenon appears.

In Figure 2.3, three horizontal lines illustrate the representative viewing distances of 0.5 m, 2 m, and 10 m, corresponding to watching content on a computer screen, a television set, and in cinema halls. The schema is suggestive for the way the human perception of depth performs differently in function of the visualization conditions (in this precise example, in function of the visualization distance). What is interesting to observe is that it is very unlikely for the

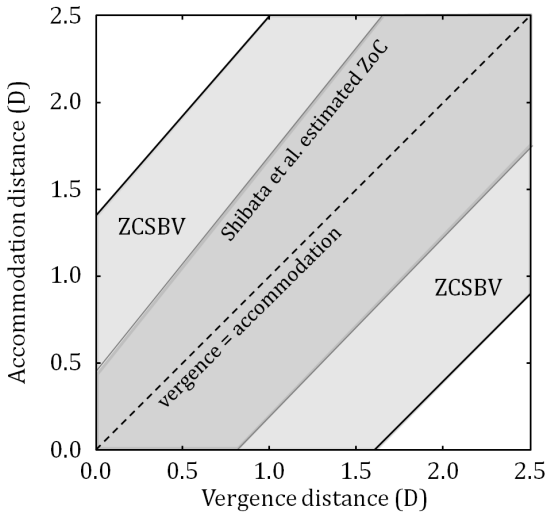


Figure 2.2: ZCSBV and ZoC estimates in units of dioptres; adapted from [Watt 2013].

accommodation-vergence rivalry to be experienced in cinema halls (except for objects represented as coming out from the screen very close to the observer), compared to a computer viewing situation, since the ZoC is wide at large distances from the screen and very small at close distances [Watt 2013].

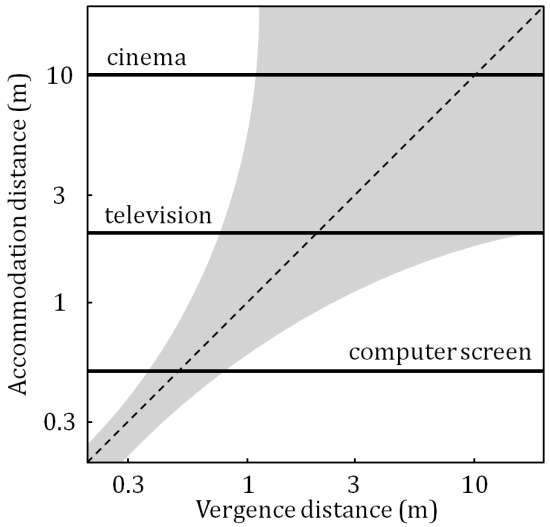


Figure 2.3: ZoC estimate in meters; adapted from [Watt 2013].

In these figures, the dioptres or the meters were more suggestive for the theoretical illustration, but, in the stereoscopic industry, the ZoC is typically expressed in percentages of screen width [Mendiburu 2009].

2.1.4 Specific vergence and accommodation manifestations

Several specific manifestations of the oculomotor system related to vergence and accommodation occur in special conditions.

The vergence state under the no-stimulus condition [Ukai 2006] is represented by *phoria*, which can be defined as a tendency of the directions of sight to deviate from normal when *binocular fusion* (Section 2.1.5) cannot be performed. There are persons for which the eyes have mostly a tendency to deviate outward (condition referred to as *exophoria*) and others for which the eyes deviate mostly inward (*esophoria*), but in all cases these tendencies vary with the viewing distance [Endrikhovski 2005].

In cases of insufficient lighting of the stereoscopic screen (less than $0.1\,\mathrm{cd/m^2}$), a particular phenomenon called *dark vergence* was identified, where the eyes have the tendency to verge at a distance of about 1 m in front of them [Patterson 2007].

Also, when no accommodative stimulus influences vision, as it is the case for dark spaces or empty fields, the accommodation of the eye lens is adjusted to an intermediate position between the far and near points. This characteristic of human vision is called *tonic accommodation* and the visual fatigue seems to influence the way it is performed at such point that sometimes, in case of considerable visual fatigue, the accommodation cannot be precisely performed for other points than the tonic accommodative point and any other accommodation induced is biased towards this point. The name given to these errors is *accommodative lead* or *lag* [Ukai 2006]. The same type of tendency to drift toward a resting distance can also be found under the name of *dark focus*. The resting distance was estimated at 75 cm from the eyes, however large individual variations exist [Patterson 2007].

2.1.5 Binocular fusion

The *binocular fusion* is performed by neuro-muscular processes. The *motoroic fusion* is the muscular process that encompasses all the vergence movements which dynamically align the optical axis of the left and right eye in order to bring corresponding retinal images into the zone of clear, single binocular vision (ZCSBV). In direct relationship with the action of the muscles, the *sensoric fusion* is the neural process of merging the two matched retinal images into a single stereoscopic image [McCarthy 2010].

The binocular fusion is possible for the points previously described as having zero disparity, however it is not limited to them. The grey zone in Figure 2.4 is the theoretical representation of the disparities that correspond to *Panum's fusion area* on the retina. This area, projected in the real-world space, represents a small region around the horopter where binocular single vision takes place for a given fixation and for which the two retinal images are fused into a single image in depth [Lambooij 2007]. In other words, it represents the total amount of disparity compatible with single vision. The existence of such a fusion area on the retina was first suggested by Panum, who considered that a retinal point in one eye could correspond to a retinal area instead of a point in the other eye [Qin 2006].

The real shape of Panum's fusion area is quite complex and only approximated as a circle in a simplified representation. A study focused on fully measuring the disparity limits of this area on the retinal fovea in 16 directions, from 0 to 360°, by steps of 22.5^o, determined that the Panum's fusion area has the shape of an ellipse off-centered toward the nasal side on the horizontal meridian of the retina [Qin 2006]. Its estimated shape is represented in Figure 2.5.

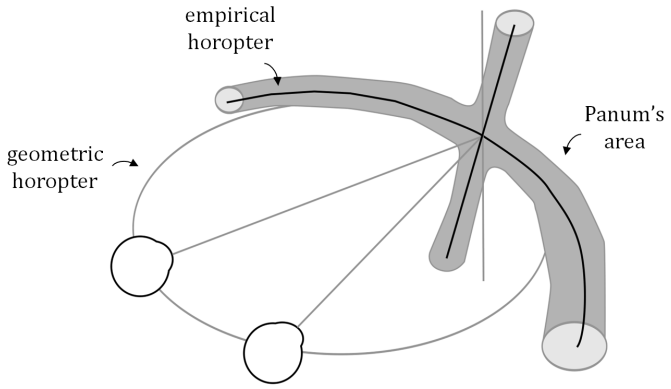


Figure 2.4: The horopter and the disparities corresponding to Panum's fusion area.

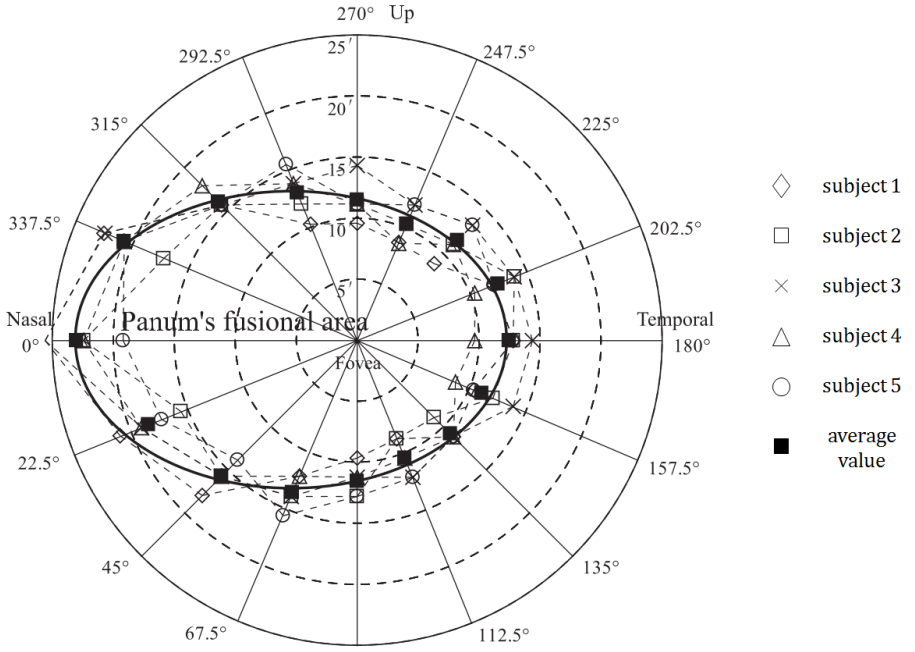


Figure 2.5: Experimental results from [Qin 2006]: disparity fusion limits for five subjects and average disparity fusion limits in 16 different directions for the right eye retina; the off-centered ellipse shown by a solid curve was estimated as Panum's fusion area in the fovea.

The *operational range of stereopsis*, that shapes the delimitation of Panum's area, can be defined through the means of two thresholds. The first one, the *depth discrimination threshold* (or *stereo-acuity*) indicates the smallest disparity differences that can be detected by the HVS. The smallest photo-receptors in the human retina have a diameter of approximately 25-30 seconds of arc, but, as the stereoscopic acuity of the eyes is as good as approximately 2 seconds of arc, the human stereo acuity is actually a form of hyper-acuity [IJsselsteijn 2002b]. For a brief stimulus duration, when there are no vergence movements, the second threshold, *i.e.* the *upper limit of fusion* or the maximum disparity that can be fused, was found to be of 27 minutes of arc for crossed disparities and 24 minutes of arc for uncrossed disparities, with the limit increasing

to up to 32 to 40 minutes of arc in the horizontal median at fovea level [Qin 2006]. In the case of longer stimulus durations, with convergence eye movements, even up to 4.93 degrees can be fused for crossed disparities and 1.57 degrees for uncrossed disparities [Yeh 1990].

Studies showed also that the HVS is not only capable of fusing horizontal disparities, but vertical disparities as well. This is natural since it was found that, if the binocular correspondence is stable, vertical disparities increase with the convergence of the eyes, due to the perspective distortion of the retinal image [Stevenson 1997]. Therefore, there is a horizontal gradient of vertical disparity across the retinal images, which depends on the distance to the stimulus and its amplitude can be regarded as a signal to viewing distance that is used to scale horizontal disparities in the interpretation of perceived depth [Watt 2013].

The operational limits for the binocular fusion capacities of the HVS in the vertical direction were found to be of up to 19 to 25 minutes of arc for simply identifiable stimuli displayed for very short time delays [Qin 2006]. However, there are situations where too large vertical disparities deteriorate the accuracy of the process of fusing the horizontal disparities. If the targets are monocularly identifiable objects, it was demonstrated that judgments from horizontal disparity are possible, even with up to 4 degrees of added vertical disparity. In tests with static random-dot stereograms however, depth judgments are impossible if even 10 minutes of arc of vertical disparity are present. Nevertheless, in tests where the random-dot stereograms are displayed dynamically, the matching sensitivity improves considerably if the number of matching elements is relatively large [Stevenson 1997].

But all these limits are not strict and the performance of the binocular fusion can vary in function of the various physical characteristics of the eyes and in function of the characteristics of the visualized content. It was shown that the stereo-acuity is better for high frequency stimuli (20 seconds of arc, compared to 5 minutes of arc for low frequency stimuli), but that the maximum disparity fused is larger at low frequencies (up to 4 degrees, compared to 40 minutes of arc at high-frequencies) [Patterson 2007]. In terms of comfort, for a high definition display, disparities of up to 35 minutes of arc were found to not cause any discomfort. However, disparities of over 70 minutes of arc are to be avoided [Pastoor 1995, Kooi 2004]. A dependency of the sensitivity to disparity on the temporal frequency of the content was determined as well, in the sense that the sensitivity to disparity requires high spatial frequency and low temporal frequency, or lower spatial frequency and moderate temporal frequency [Patterson 2007].

The eccentricity of the impressed retinal area has also an important influence on the perception of disparities, where the *eccentricity* can be defined as the angular distance from the center of the visual field or from the fovea of the retina. This happens because the distribution of the various vision cells on the retina is different for different eccentricities. The size of the receptive fields increases and their spatial frequency decreases with retinal eccentricity. As a consequence, spatial acuity is highest at the fovea, falling off rapidly toward the periphery [Liu 2010]. This justifies the fact that the HVS is less sensitive to the peripheral areas of sight [Gorley 2008].

Thus, while the depth discrimination is best at the foveal center and for zero disparities, and decreases with growing eccentricity and for non-zero disparities, the upper fusion limit increases with eccentricity and stimulus size [Liu 2008]. As a consequence, with eccentricity growing, the depth discrimination is less precise for objects located at different depths than the fixated point. Therefore, the shape of Panum's fusion area not only greatly depends on eccentricity, but it also depends on the characteristics of the fixed point and on the surrounding stimuli [Liu 2010].

A strong correlation has also been found between the limits of binocular fusion and the depth of field. The explication comes from the fact that beyond the DOF limits, blur is present, which makes the Panum's fusion area expand, in the sense that binocular fusion will be performed in those regions of low spatial frequency and disparities will no longer be distinguished. Thus, the limits for the capacity of distinguishing disparities are very close to the limits of clear sight in an ideal theoretical situation, when the other multiple factors that influence the two are not taken into consideration [Lambooij 2007].

Also, the total number of neurons sensitive to near zero disparities is larger than the total number of neurons sensitive to stimuli nearer or farther than the fixation. Further more, all the vision cells have ranges of *preferred disparities* (limited to ± 1 degree [Liu 2008]), which explains for example the difficulty of the HVS to fuse disparities for objects located too near.

Another particularity of the HVS is the fact that the number of visual cells tuned to perceive near disparities is slightly larger than the number of visual cells that capture far disparities and better stereo-acuity was reported for crossed than for uncrossed disparities [Liu 2008].

In cases where binocular fusion cannot be performed, the phenomenon called *diplopia* (or *double vision*) appears, which is the condition where the left and right corresponding images remain separate instead of being fused into a single image [McCarthy 2010]. As a general rule, the HVS is more sensitive to stereo mismatches in the vertical dimension than in the horizontal dimension [Kooi 2004].

2.1.6 Binocular rivalry

Particular manifestations of the HVS occur in the case when different content is shown through a stereoscopic system to the two eyes or when the stereoscopic content displayed contains too large disparities, an important misalignment, or other image distortions [Patterson 2007]. In this case, the left and the right views cannot be fused and *binocular rivalry* produces. Multiple reactions of the visual system have been observed, depending on the type of stimuli presented to each eye.

In situations when corresponding parts of the two retinas receive very different high contrast images that cannot be fused, *binocular suppression* occurs. This is a phenomenon opposed to *binocular summation*, which is typical for cases of correct binocular fusion. The consequence of binocular suppression is that each image is perceived alternately in reciprocal periods, in the sense that at a given moment only one of the two images (the so-called dominant stimulus at that specific moment) is perceived, while the other is temporarily suppressed [Boev 2010, IJsselsteijn 2002b].

When there are mismatches of luminance or color in areas of uniform illumination, the brain successfully effectuates the binocular fusion, but a certain shimmer is also perceived, which makes the area impossible to be localized in depth [IJsselsteijn 2002b]. This behavior of the HVS is called *binocular luster*.

Binocular mixture is also a phenomenon which occurs in situations where the content delivered to each retina is very different. But its typical manifestation is in cases where one eye sees an uniform field and the other sees a detailed stimulus [IJsselsteijn 2002b].

Another observed behavior in the cases where binocular rivalry influences the perception is that *high spatial frequency content* generally dominates the *low spatial frequency content* [Lu 2009].

Some of these processes that are typical to binocular vision illustrate for example how human observers seem more tolerant to coding distortions in stereo than in non-stereo sequences, given the fact that the stereoscopic perception often effectuates compensations [IJsselsteijn 2002b].

2.1.7 Safety and health issues

On the one hand, various dysfunctions of the visual system can prevent people from correctly perceiving stereoscopic depth. These can even manifest as *stereoblindness*, which is the complete incapacity of perceiving the binocular depth cues.

A frequent such unhealthy eye condition is for example *strabismus*, which, as a consequence of the improper coordination of the eye muscles, causes imperfect binocular vision and difficulties in binocular depth perception. But, with about 6-8 % of the population touched [Patterson 2007], strabismum is just one example.

Statistically, it was found that approximately 20 % of the population suffer from some form of binocular anomaly [Lambooij 2009]. For example, one explanation for binocular anomalies is related to problems in the functioning of cortical neurons and there is no known remedy for such individuals who lack the neural capacity for binocular vision [Patterson 2007].

On the other hand, the long exposure to 3D content on stereoscopic displays may lead to a deterioration of the visual health condition of the viewers, especially in the case where the 3D content is not of good quality, *i.e.* when it contains excessive impairments or excessive depth effects. An example to support this fact is the experimental proof of a decline of visual functions caused by the accommodative vergence stimulated when watching 3D content with heavy vergence loads [Emoto 2005].

When frequent, abrupt or extreme depth discrepancies are present in a stereoscopic 3D video, the discomfort felt by the viewer is represented by the psychological and perceptual condition called *depth change stress* [McCarthy 2010]. This could be the case of the imperfect scene cuts in some 3D movies. For a comfortable viewing, these cuts should optimally prove a smooth depth transition.

Observers can also feel discomfort when watching stereoscopic images with a dense content and multiple points of interest. Since there is no obvious point of reference, the muscular and neural effort of the HVS is considerable, because of excessively moving the gaze from one point to another, between different depths. The situation is being referred to as *fixation point conflict* or *reference point conflict* [McCarthy 2010].

The *simulator sickness* notion is inspired from the world of flying simulators and is often used in the field of stereoscopic quality perception to describe the effects of the conflicts between the visual perception and the vestibular system, when motion is involved. Because of this type of conflicts, motion can be incorrectly perceived by viewers. More on the simulator sickness symptoms will be discussed later, in Section 4.7.6.

Since *the visual system of children* is not fully developed until the age of seven, some ophthalmologists believe that viewing stereoscopic content could harm them by generating strabismus. A study on this subject from 1988 proved that stereoscopic content influenced the binocular visual system of a four-year old child, who manifested esotropia (a form of strabismus) as a consequence. But this was an isolated study and no other results exist that confirm this hypothesis or deny it. Therefore, in lack of more evidence, special attention needs to be paid to the way stereoscopic 3D content is viewed by children [Ukai 2006].

What is certain is that many of the properties of the human vision change in time, differently for each person. While the vergence ability of the two eyes generally remains unchanged, the transmissiveness and the accommodative power of the eye lens diminish with age [Ukai 2006]. This affects multiple perception abilities of the eyes, like the sensitivity to glare, the color sensitivity and, broadly, the binocular depth perception. There are also retinal and neuronal changes that usually occur after the age of 55 and these can influence in a negative way the size of the visual field or the sensitivity to low light levels [IJsselsteijn 2002b].

2.2 Mechanisms of human depth perception

Based on the physical properties of the HVS presented in the previous sections, the human perception and understanding of depth is a complex mechanism triggered by the interpretation of various *depth cues*. These are visual or physiological elements that give us hints about the localization of objects in space (i.e. the *layout* [Cutting 1995]) and they can be referred to as *binocular* or *monocular*, depending on whether they can suggest three-dimensionality when using both eyes or even when using only one eye.

2.2.1 Binocular cues

Binocular disparity (on the retinas) or *binocular parallax* (on the stereoscopic screen) is considered to be one of the strongest depth cues for the HVS [Boev 2010]. The *disparities* help create the sensation of depth when the two retinal images are combined by neuronal processes. However, the variation of the disparities in the real world is different than the way disparities vary on a stereoscopic screen, therefore, when watching 3D content on a display, this cue can sometimes be ambiguous in defining the exact depths in the scene [Patterson 2007].

Another binocular mechanism that offers information related to the depth of an object in space is the *vergence* of the eyes when looking at that specific object. Thus, the angle formed by the two axes of view of the eyes indicates the position of the targeted object.

2.2.2 Monocular cues

The *accommodation* of the eye lens (also referred to as *defocus* [Torralba 2002]), an *oculomotor cue* (or *physiological cue* [Lipton 1982]) that is related to the feedback of the muscles that control the eyes [Holliman 2006], can also give indications on the distance to an object. It is the primary depth cue for very short distances, where an object is hardly visible with two eyes and

its importance relative to other cues quickly decreases when the distance to the object increases [Boev 2010].

Other monocular cues are *occlusion* (or *interposition*), *light and shade*, *linear perspective*, *relative size of objects*, *relative density*, *textural gradient*, *atmosphere blur and saturation*, *depth of field*, and *the previous knowledge of shapes ans sizes* [Cutting 1995, Holliman 2006, Mendiburu 2009, Devernay 2010]. They are also called *pictorial cues*, being often used in two-dimensional representations for suggesting depth. Some of these monocular depth cues are illustrated in Figure 2.6.

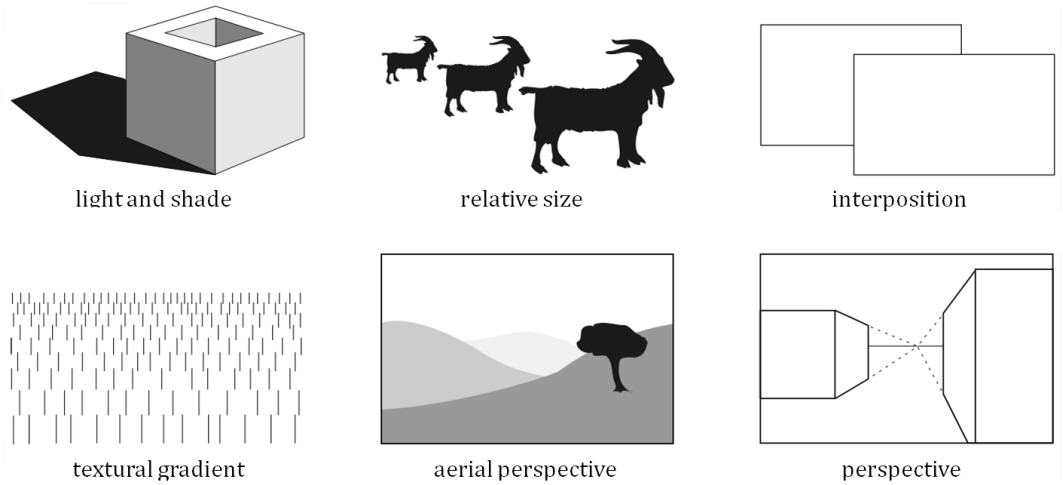


Figure 2.6: The illustration of several pictorial cues [Devernay 2010].

All these can be considered *experiential depth cues* as, over time, observers learn the physical significance of different retinal images and their relation to objects in the real world [Holliman 2006]. *Occlusion* is, in almost all cases, dominant over all the other cues [Lambooij 2007].

Motion parallax appears when an observer moves his eyes or head. The fixation point stays automatically on a specific point and the closer and further images move relatively to each other on the observer's retina [Boev 2010]. This relative movement gives the observer hints on the real depths. Mendiburu elaborates more on motion parallax, separating this depth cue in two different mechanisms: *parallax induced by point-of-view movement* and *parallax induced by object's movement* [Mendiburu 2009].

Motion parallax represents a very strong depth cue, along with *binocular disparity*, and the two combined lead to a better depth perception than that inferred by either cue alone [Holliman 2006]. It was stated that the two cues are strongly related, since temporally separated successive views can provide the same type of information to the visual system as spatially separated views [Rogers 1982].

It is considered that *motion parallax*, *binocular disparity*, and *defocus* are cues for absolute depth measurements. In their absence, the absolute distance between the observer and a scene cannot be measured by the HVS [Torralba 2002].

2.2.3 Integrating the depth information

For an accurate three-dimensional perception, the human visual system needs to integrate the information extracted from each cue that is present in a specific context.

In a simplified representation, the integration of the various cues can be regarded as a *linear combination*, in which the weight of each cue is proportional to its relevance in the context. However, "the *cue weights* cannot be fixed values, or learned for certain classes of situations, because the relative reliabilities of depth information from different cues depend on the particular viewing situation, and may vary substantially even within a simple scene" [Watt 2013].

Cutting *et al.*, convinced that an exhaustive combinatorial exploration of the interactions between the existing depth cues is impossible because of their large number, proposed an individual study of several cues and of their efficacy at different distances, "in an attempt to prune the apparent richness of information about layout at all distances to a more manageable arrangement within particularly domains" [Cutting 1995]. Their conclusions were synthesized in a suggestive graphical representation, which is structured according to their classification of the types of space around an observer (*personal space*, *action space*, and *vista space*). This representation is reproduced in Figure 2.7, in an adapted form [IJsselsteijn 2002b].

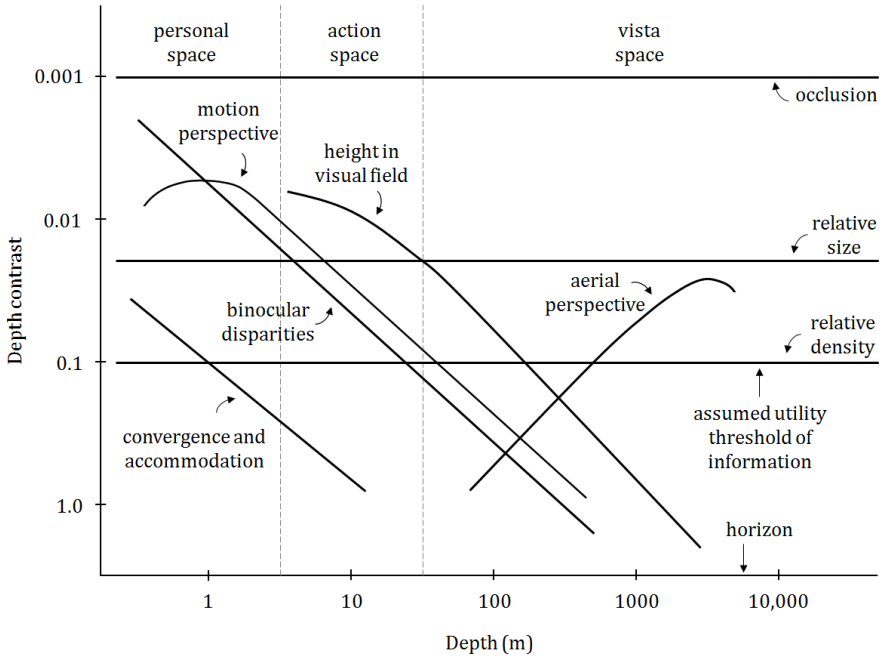


Figure 2.7: The relative strength of several depth cues at different distances [Cutting 1995].

Similarly, a simplified visual representation of the depth cue integration by the HVS in function of the distance from the observer to the scene was proposed by Boev *et al.* [Boev 2010]. The adapted schema is presented in Figure 2.8.

Besides the simple presence of a certain depth cue, its *reliability* influences the role that this cue plays in the perception of depth. The reliability of a depth cue, according to Jacobs, can be determined either by the amount of ambiguity of the cue, or by its consistency with the other cues [Jacobs 2002]. The HVS has thus the difficult task of having to identify and integrate

signals that refer to the same object and avoid integrating the signals that do not [Watt 2013].

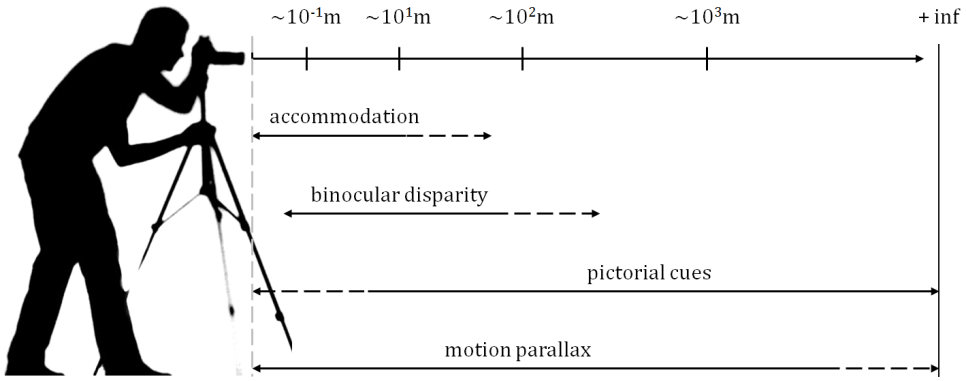


Figure 2.8: Depth perception viewed as the integration of a set of *layers* of depth cues, varying with the distance to the scene; adapted from [Boev 2010].

Thus, very complex ways of combining the visual information can be observed in the behavior of the HVS. Howard et al. propose explanations through mechanism like: cue averaging and summation, cue dominance, cue dissociation, cue reinterpretation, or cue disambiguation [Howard 1995].

As a result of an experimental study on the depth cues interaction, Young *et al.* stated that, in the case of depth vision with cues strongly in conflict, a full description of the depth combination rules for inconsistent and impoverished stimuli seems likely to resemble a microcosm of cognitive processing: elements of memory, learning, reasoning and heuristic strategy may dominate [Young 1993].

To illustrate the complex interference between the various cues, an example is the fact that the texture information and the binocular information have different influences on the overall depth perception of a textured object in function of the slant of that object. It has been proved through experiments that texture is very reliable at high slants but unreliable at low slants, when the stereo information comes to compensate for a good depth perception [Knill 2003].

Despite the complicated correlations between the cues, a general argument is that the more consistent the cues are, the better the three-dimensional perception is [IJsselsteijn 2005]. Furthermore, monocular cues are perceived independently of binocular cues and Julesz has proved, through random-dot stereogram experiments, that the binocular disparity information can be sufficient for perceiving depth, even when all the other cues are missing [Julesz 1971].

To conclude, although some researchers seems to show consensus on depth cue integration in terms of *statistical inference*, the discussions regarding the strategies used by the brain to extract depth are still open and no single unified theory about cue integration has been established so far [Lambooij 2007].

Conclusions

This chapter offered the key concepts on which the work presented in this manuscript was based.

It highlighted the main *physical characteristics and mechanisms* of the HVS that allow the perception of the surrounding world in *three dimensions*. It also mentioned the *limitations* of the *HVS* or how these limitations may change in time or in function of exterior stimuli.

The chapter explained as well the situations in which the visual perception of the depth can be *faulty* or when the stimuli watched can *harm*. It introduced the various *cues* that the brain uses in order to interpret the various visual stimuli as objects *in depth* and discussed various theories on their integration.

What is certain however is that numerous human vision processes are not yet fully understood and that the HVS remains a fascinating universe to explore.

Chapter 3

The Stereoscopic Systems

Introduction

Regardless of the numerous improvements brought to them in time, the devices and the techniques needed for the manipulation of *stereoscopic content* are still complex and quite complicated to use. The difficulty of exploitation is mostly due to the *tight interconnections* among the various components of a stereoscopic system, which impose a precise control of all the parameters at every step of the manipulation chain.

A correct insight on the main principles of *producing*, *storing*, and *rendering* of the stereoscopic data is thus essential when having in view the understanding of the factors that influence its quality. The current chapter contains precisely such a presentation.

3.1 Dual-camera geometry

For recording stereoscopic content, two cameras are necessary, in order to simulate the two angles of view of the eyes. The two cameras need to be placed at different positions on a horizontal plane and the left camera is supposed to record the images to be delivered to the left eye and the right camera the images to be delivered to the right eye. The distance between the two cameras is called *camera-base distance*, *camera-base separation*, or simply *baseline*.

Two main types of configurations for the *dual-cameras* exist. One is the *parallel configuration*, in which the optical axes of the two cameras are placed on parallel directions and the other is the *toed-in configuration*, in which the camera axes are converging. From the two, the parallel arrangement is generally preferred, since it introduces less geometrical distortions. The toed-in configuration is even suggested to be avoided when possible [IJsselsteijn 2000], because the keystone distortion it inevitably produces (aspect discussed in Section 4.5.1) can be very difficult to correct.

Figure 3.1 illustrates the two types of stereoscopic camera configurations: two toed-in configurations with different convergence angles to the left and two parallel configurations with different baselines to the right.

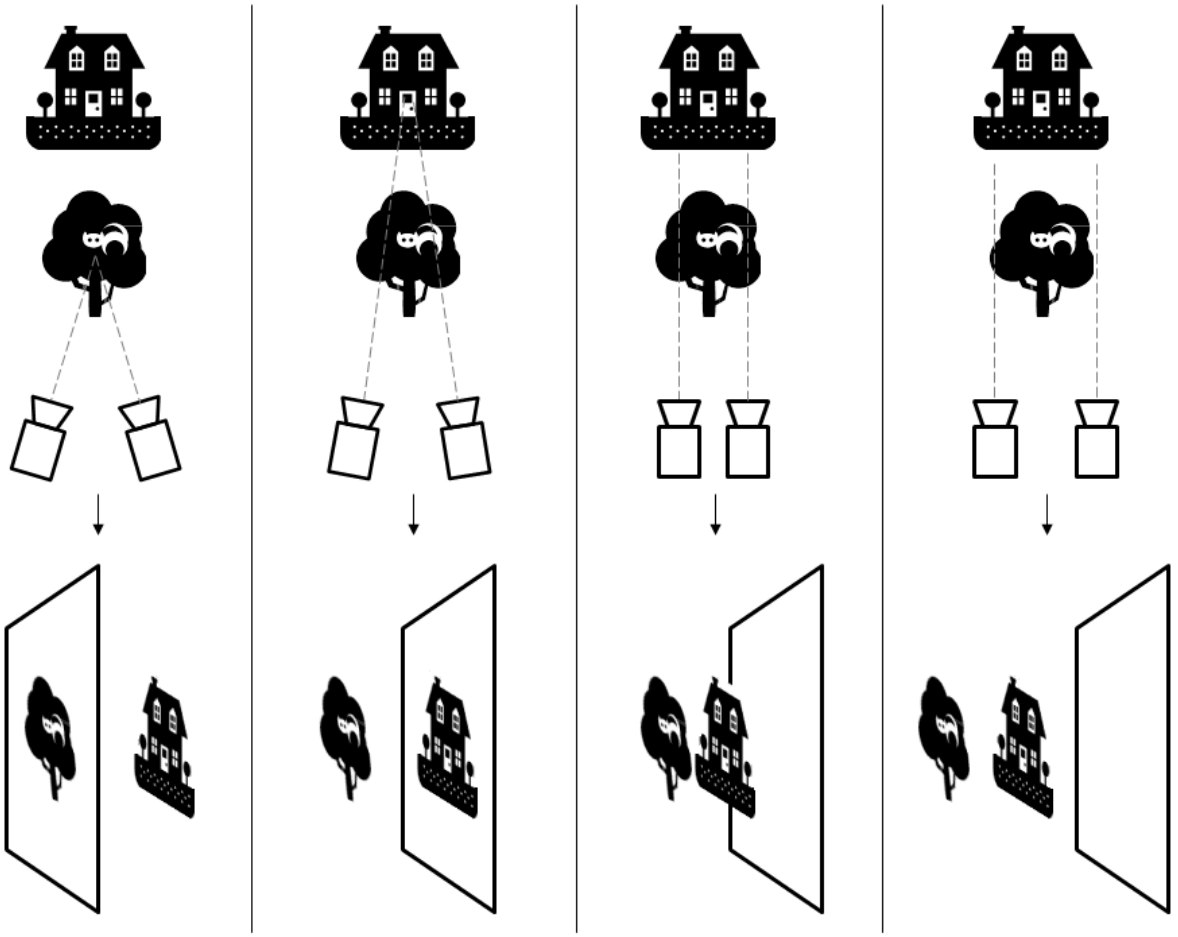


Figure 3.1: The toed-in and parallel camera configurations and the variations in the rendered depth relative to the screen plane in function of the dual-camera convergence angle and of the camera-base distance; adapted from [Mendiburu 2009] and [Sony 3D Television & Film 2010].

The same figure also explains the way the rendered depth is dependent on the camera-base distance and on the convergence angle between the two cameras at capture. On the one hand, it illustrates that the *recorded disparity* is directly proportional to the baseline (*i.e.* a larger baseline generates larger depths) and, on the other hand, that the *convergence point* of the dual-cameras determines the plane to be represented at screen level during the visualization,

with the objects in front of this point represented in front of the screen and with the objects behind this point represented behind the screen.

The recorded disparities are also influenced by the focal length of the cameras and by the distance from the cameras to the fixated object: the larger the focal distance, the larger the recorded disparities and the larger the distance to the object, the smaller the recorded disparities [IJsselsteijn 2002b].

3.2 Dual-camera capture

During the capture of stereoscopic images, a particular attention needs to be paid to the recording parameters.

A precise *calibration* needs to be performed for both cameras in terms of color, brightness, contrast, focus, or geometry [IJsselsteijn 2002b]. In the mean time, *no vertical misalignment* is accepted between the two cameras, *nor any rotation* in the vertical plane that comprises the two cameras. These constraints are a consequence of the properties of the HVS, which is disturbed by any mismatch that can exists in the two views of a stereoscopic pair, with the only exception of the horizontal disparities that it translates into the impression of depth.

The envisaged display size and visualization distance must be taken into consideration as well and the recording parameters must be adjusted accordingly. This is necessary because there is a *strong link between capture and rendering*, and the content recorded with a certain set of parameters will not be suitable for rendering on any type of display, as it will later be shown in Section 3.6.

There are many mathematical aspects related to the stereoscopic parameters with which the camera operators must be familiar in order to produce good quality recordings that do not harm the visual system of the viewer.

An example of a very simple rule that appears to be successfully used by amateur stereoscopic photographers for producing correct content for their home displays is the *1/30 rule*. This rule says that the baseline should be 1/30 of the distance from the dual-camera to the object to be captured and this ratio serves to correctly adjusting the baseline in function of the distance to the subject, or vice versa, such that the maximum screen disparities would not be too large for a comfortable viewing.

A large number of books exist however on numerous more elaborated principles to be followed when producing stereoscopic content [Spottiswoode 1952, Lipton 1982, Lipton 1997, Mendiburu 2009, Devernay 2010].

3.3 3D content compression

Once stereoscopic data is produced, storing comes into play. In most cases, stereoscopic 3D data, as it is captured, represents the double of the equivalent 2D data, because of the necessity of supplying content for both eyes. Therefore, compressing the stereoscopic data before storing it becomes a necessity.

3.3.1 3D data redundancy

The encoding of visual information in general can be approached in a variety of ways, due to the different types of redundancies that it usually contains.

Correlations that can be exploited exist *between adjacent pixels of the same image* and, in the case of video data, between pixels of the same frame (*intra-frame redundancy*) and between corresponding pixels of successive frames (*inter-frame redundancy*). Thus, methods for predicting the values of certain pixels from others surrounding them can be used, leading to a reduction of the final size of the file on disk.

By taking into account the properties of the HVS, further consistent compression can be achieved for image or video data sets. This is possible first of all because the human eye has a limited power of distinguishing between very similar colors and shades and the very large range of colors and shades that exist in reality can thus be reduced to a very small scale during encoding, without the viewer perceiving a lack of details. Another important factor that can contribute to efficient compression is the different sensitivity of the HVS to different types of stimuli. For example, the human sight is more sensitive to changes in areas of low frequency content, therefore, a more significant compression can be applied to areas of high frequency, without loss in the quality of perception [Boev 2008]. Classical algorithms of *entropy coding* that can be used for this purpose are for example the Huffman coding or the arithmetic coding.

A special case of compression is that of the stereoscopic content, in which there are many similarities between the left and right images or between the left and right video sequences (*inter-view redundancy*).

When compressing stereoscopic content, due to the various interactions between the perception of the two views (binocular suppression, binocular summation, binocular mixture), a high potential for reducing the stored data exists. Applying a different compression level to the two views has proved efficient for optimizing the size of storage and consistent research is undergone that is oriented towards refining the method and defining the boundaries between which the quality of perception is maintained [Seuntiens 2006b, Lu 2009, Aflaki 2010b, Gorley 2010].

3.3.2 3D data compression principles

The main classification of any compression scheme is as *lossless* or *lossy*. The insertion of a scheme into one of these two categories depends on whether the exact initial content can be recovered from the data after compression and decompression. And, among the lossy schemes, especially when speaking of compression for visual content, a particular type is that of the *perceptually lossless* compression schemes, which discard significant parts of the original content, though without diminishing the perceived quality.

Multiple manners of compressing stereoscopic data exist. Most of them can be included in one of the following three categories: disparity based methods, methods based on theories of binocular suppression, and depth map coding methods [IJsselsteijn 2002b, Meesters 2003].

Since the two views that make up the stereoscopic content are very similar, a good method of reducing size is to store only one of the two views (the one chosen as a *reference view*), along

with the *disparity information* that can help restore the second view (the so-called *target view*). This is a *disparity based approach* and its main problem is that often by storing only one of the two views, when trying to recreate the second view, not enough information is available, and the occluded areas in the reference view cannot be accurately reproduced in the target view. Thus, artifacts are formed. To avoid this problem, a solution could be the inclusion of supplementary original data in the compressed set.

Stereoscopic data can also be stored as it is, made up of the two different views. In this case, consistent compression can be achieved by coding the two views differently, following principles of the binocular perception of the HVS. This type of compression can be implemented as *asymmetric coding* (or *mixed resolution coding*), and its main principle is to compress one of the two components of the stereoscopic pair more, either by reducing its quality, or, respectively, by reducing its resolution. One view is meant to represent the details of the scene, while the other view is compressed in such a way that it only contributes with the disparity information [Gorley 2010].

Another type of compression for 3D is employed when the depth information is stored separately from the 2D data from the beginning. This happens for example in the case of special 3D cameras that store a 2D image and its associated depth map. The compression is implemented in this case as a *depth map coding method* and is performed for both the 2D image and the depth map. Improvements to the method can be brought if *regions of interest* (ROI) are identified in the visual information. Thus, the compression can be stronger in all the areas that are not regions of interest [Meesters 2003]. For this type of content made up of 2D data and depth, disocclusion (discussed in Section 4.5.2) is as well a significant drawback.

There are also methods for coding 3D data for multi-view systems, where precise views of the 3D objects from several viewer position are intended to be obtained. This is, for example, the case of certain auto-stereoscopic displays with multiple fixed views or of certain displays with head-tracking capacities, both of which change the displayed images in function of the viewer position, in order to maintain natural perspectives. 360 °-view displays also exist that need compression techniques for their very large amount of data. The *multi-view coding methods* for such data have a considerable potential for compression by exploiting the inter-view redundancies. Their main approach is also disparity based, since the 3D data can be recreated by rendering any desired view from a set of key views and the associated disparity information.

The suitable compression scheme is always to be chosen on a one-to-one basis according to the context - the capture device, the recorded content, the technology to be used for visualization, compatibility issues, *etc.* And, in order to assess the performances of several compression schemes for a given content, a simple and intuitive method is to compare their achieved size reduction and also the perceived quality of the compressed content [IJsselsteijn 2002b].

3.4 3D data formats

With the ever increasing interest in stereoscopy in numerous contexts, standardization is now imperative. New media formats are being developed, which should satisfy the need of compatibility and interoperability between different devices.

These new formats also need to satisfy the condition of backward compatibility with existing

technologies. This is a necessary condition for the spread of stereoscopic content on a large scale [Smolic 2009].

As a consequence, a simple solution is to store stereoscopic content in common existing formats, like JPEG for example, but with the compromise of compressing the two components of the stereoscopic pair. This can be done with a loss of half the horizontal resolution, in the case of the *side-by-side* method, or with a loss of half the vertical resolution, in the case of the *top-bottom* method.

There are however specific formats that already manage to store the full-resolution 3D data. We present here a single example, that of the *Multi-Picture Format*, abbreviated as *MPO*, which is able to store two or more JPEG files concatenated together, along with other additional information specific to stereoscopic data or to multi-view data.

3.5 3D displays

Two major categories of existing stereoscopic 3D displays can be differentiated by the way they deliver the depth information. The first are classical (or very similar to classical) displays that use additional devices for performing the left-right separation (*i.e.* for delivering the left and right images displayed on the screen correspondingly to each eye). Most frequently, the additional *separation devices* are *special eyeglasses*. The second category of devices is represented by those screens that deliver the correct stereoscopic content directly to the eyes of the observer(s). These are called *auto-stereoscopic displays*, while the first category of displays is often referred to as simply *stereoscopic displays*.

The separation of the left and right components displayed on a stereoscopic screen can be *time-parallel* or *time-sequential*. As the names suggest, in the time-parallel case all the stereo information is displayed simultaneously on the screen and in the case of time-sequential displays the left and right content is displayed alternately in time.

3.5.1 Anaglyph systems

One of the oldest types of time-parallel stereo separation is *color multiplexing*, detailed in Figure 3.2. This is a classical stereoscopic approach in which both the left (1) and the right (2) channels of information are superposed simultaneously on a normal screen (5) after having been previously filtered with complementary colors (3 and 4). With the means of *anaglyph glasses* (6), which have lenses of the same two colors used for the image filtering, the redundant content is separated (7 and 8) and sent correctly to each eye.

The main colors used for this kind of stereoscopic separation are: red-cyan, yellow-blue, and green-purple. Examples of anaglyph glasses using them are shown in Figure 3.3. Among these, the red-cyan ones are the historical first type of anaglyph glasses used in the past.

The color multiplexing method is very cheap to implement, since every color screen can display the superposed left and right filtered views, while the cost of the glasses themselves is very low. However, this separation technique is not very precise and often produces very poor color performances, like color skewing or unsaturated colors. Also, the color compression

is almost impossible. Therefore, although the anaglyph glasses have been the main separation technique used in cinemas in the past, at present they are only interesting for personal utilization in the home, where they can allow occasionally viewing short 3D sequences for a minimum cost, when other more expensive or sophisticated technologies are missing [Borel 2013].

Figure 3.2: The functioning of a stereoscopic system using anaglyph glasses; adapted from [Sony 3D Television & Film 2010].

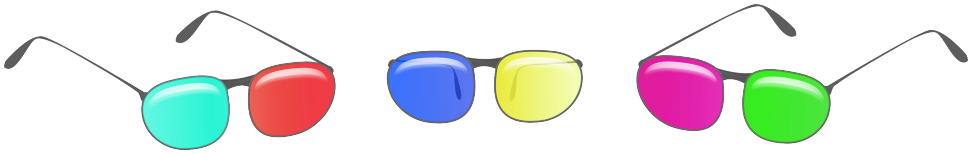
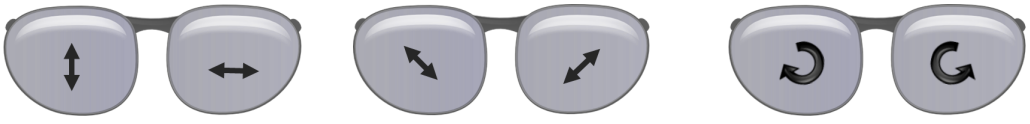


Figure 3.3: Different types of anaglyph glasses.

3.5.2 Polarization systems

A much more precise way of displaying stereoscopic content in a time-parallel manner is by *polarization multiplexing*. In this case, the two different left and right images are simultaneously displayed on screen, by interleaving their pixel components in a previously selected fashion and by selectively polarizing the left and right content in a different way. In correlation with a pair of polarized glasses, which use the same polarization as that applied to the left and the right content on the screen, the correct images are decoded and delivered to each eye.

The polarization of the light can be linear or circular, with better performances in the case of the circular one, since artifacts can thus be avoided when the observer moves his head from the perfect horizontal position. Figure 3.4 exemplifies how the two types of polarization are implemented in the glasses.



(a) Linear polarization. (b) Circular polarization.

Figure 3.4: Types of polarization glasses; adapted from [Sony 3D Television & Film 2010].

The implementation of such a system imposes therefore the presence of polarization filters at two levels: in front of the screen and in the glasses. As a consequence, despite the reduced

cost of the polarized glasses (often made of plastic), the polarization systems impose an added cost compared to the anaglyph systems and this is due to the necessity to adapt the screen [Borel 2013].

Also, the resolution of the left and right views is reduced by the fact that both views are displayed simultaneously and have to fit together on the screen. Most often, the separation on screen is done by vertical interleaving, therefore there is a loss of half of the vertical resolution.

But polarization systems with no resolution loss exist as well. The differently polarized left and right views can be displayed on screen alternately very fast, while the glasses perform the separation. The method proves the great advantage of reduced crosstalk (artifact discussed in Section 4.5.4) [Meesters 2003], but decreases even more the light efficiency of the system, which is already feeble.

Figure 3.5 schematizes the functioning of a polarization system: the original left (1) and right (2) views, the same images after complementary linear or circular polarization and at half resolution (3 and 4), the two views combined on screen (5), the polarization glasses (6), and the decoded left (7) and right (8) views, again at half resolution for each eye.

Figure 3.5: The functioning of a polarization stereoscopic system; adapted from [Sony 3D Television & Film 2010].

Even if they show several inconveniences, the polarized systems are appealing in multiple contexts. In cinema halls for example, despite the initial investment in a special *silver screen*, which is polarization-preserving, the use of the cheap and sufficiently precise polarized glasses is very convenient. In the home, the polarized systems are attractive as well, due to the simplicity of using the glasses (lightweight, no battery needed) and to the affordable prices at which they can be bought [Borel 2013].

3.5.3 Modern adaptations of the classical stereoscopes

Examples of the time-parallel separation principle are also the famous *stereoscopes* proposed by Wheatstone and Brewster. They are perfect illustrations of *location multiplexing*, as the left and right content is placed adjacently in front of the user. The 3D effect is perceived on a third image, obtained after the superposition of the pair of stereo images with the help of the stereoscope. The device has a separation surface between the eyes and special lenses that help create the fusion with less effort than if trying to superpose them exclusively with vergence and accommodation mechanisms during free-viewing. The stereoscope is probably the best known and certainly the oldest method of displaying stereoscopic images [IJsselsteijn 2002b].

Inspired from the Wheatstone mirror stereoscope, there also exist some modern stereoscopic displays that use the same principle, but replace the static images with images on LCD[1] supports or other types of displays [Holliman 2006].

3.5.4 Shutter systems

The stereoscopic displays with probably the best performances today are the time-sequential displays that use *shutter glasses*. As it is shown in Figure 3.6, the left (1) and right (2) eye content is displayed on the screen alternately for very short sequences of time (3) and the shutter glasses (4) block the view for each eye in a synchronized manner, such that, at a certain moment, the correct view is seen by one eye, while the other is completely blocked (5).

Figure 3.6: The functioning of a stereoscopic system using shutter glasses; adapted from [Sony 3D Television & Film 2010].

Since the speed of the left-right views alternation is very fast, the viewer does not perceive the feeling of blocked eye-sight. On the contrary, due to the HVS property of integrating a stereo pair across a time lag of 50 ms [Meesters 2003], it perceives a fluid view of the fused stereoscopic content.

The systems using shutter-glasses have the great advantage of displaying full-resolution images to both eyes. Still, a slight decrease in luminosity generally occurs in comparison to watching the screen with no shuttering glasses, because of the fact that the lenses of these special shutter glasses block each eye half of the time [Meesters 2003].

From the implementation point of view, the technology in such systems is more sophisticated, requiring displays with elevated refresh rates (minimum 120 Hz) and an infrared emitter for performing the synchronization between the glasses and the screen (better synchronization in the case of plasma screens). The shutter glasses are expensive as well, bulkier than the anaglyph or the polarized ones, less comfortable to wear, and dependent on the batteries that need to be changed periodically [Borel 2013].

Consequently, for the quality of the experience, the shutter systems are often adopted for utilization in the home, where the maintenance of several pairs of such glasses is not cumbersome. However, their use in cinema halls is complicated, implying a considerable managing cost, therefore the number of cinemas who adopted such systems is very small.

[1] liquid-crystal display

3.5.5 Auto-stereoscopic displays

The *auto-stereoscopic displays* implement advanced technologies to deliver the correct image to each eye without the help of any additional devices. The existence of such comfortable-to-use screens is a very important step in removing an obstacle to the acceptance of 3D displays for everyday use [Holliman 2006].

Today, the two most successful technologies used in the auto-stereoscopic displays are the parallax barriers and the lenticular systems, both being time-parallel location multiplexing approaches.

The *parallax barrier displays* are an adaptation of classical displays by adding a special layer made of black strips situated at specific positions over the pixels. Since the pixels of the left and right images are aligned on the display in a precise pattern (usually interlaced columns), the barrier will block the back-light of the display in such a way that, from a specific viewing position, only the left image pixels will be seen by the left eye and only the right image pixels by the right eye [Holliman 2006].

This principle is illustrated in Figure 3.7, where it can also be noticed that this type of systems ensure an optimal viewing only for very precise positions of the observer, called *sweet spots* or *viewing windows*. The notations in this schema represent:

e: the viewing window width;

i: the pixel pitch;

b: the barrier pitch;

g: the gap between barrier and pixels;

z: the optimal visualization distance.

Another particular way of implementing the parallax barrier is also possible by using mobile slits as a barrier. This would mean switching from location multiplexing to time multiplexing [Meesters 2003].

The *lenticular system displays* are also an adaptation of classical displays, this time by adding in front of the pixels a layer of cylindrical columns of lenses called *lens arrays*, specially positioned in correspondence with the column interleaving of the left and right image components. Due to their refractive properties, these columns of lenses ensure the proper distribution of left and right information directly to the correct eye [Holliman 2006].

The principle of functioning of such systems is described by Figure 3.8, where the notations are the same as for Figure 3.7, except for l, which is the lenticular pitch, and for f, which is the focal length of the lenticular elements.

In the schema, we can also notice that typically the viewing window width is taken to be the average eye separation, $e = 65\,\text{mm}$, to give some freedom of movement (up to $e/2$) around the nominal viewing position [Holliman 2006].

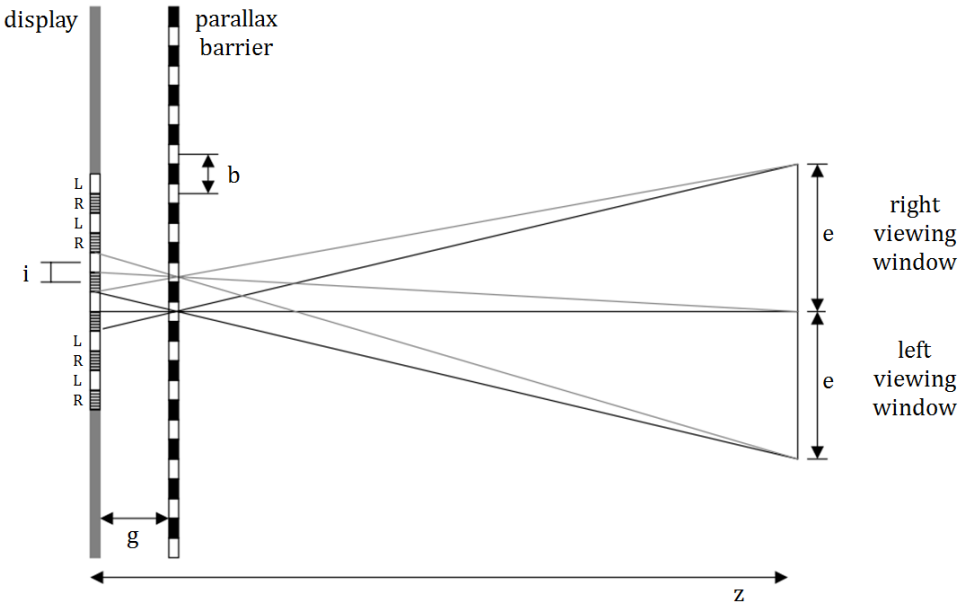


Figure 3.7: The functioning of a parallax barrier [Holliman 2006].

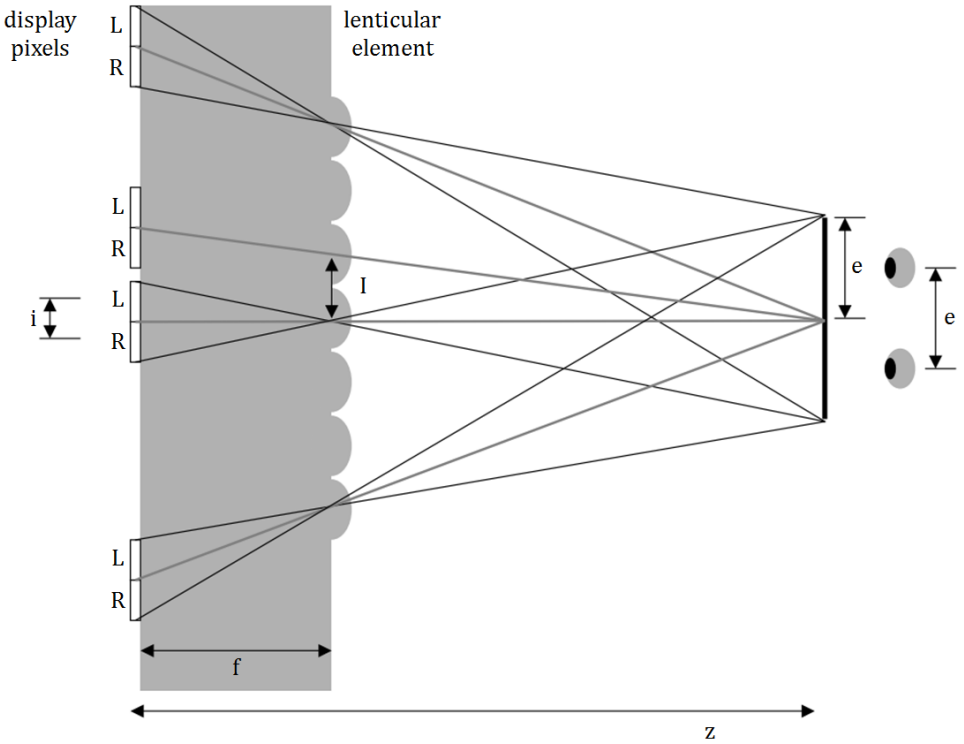


Figure 3.8: The functioning of a lenticular system [Holliman 2006].

Overall, the lenticular displays have proved better results than the parallax barrier displays, more precisely, better quality content with less artifacts [Meesters 2003]. The cheaper technology, the parallax barrier, has though the great advantage that it can be switched off anytime,

allowing for backward compatibility of the display with 2D content. But, it also has the disadvantage of a decrease in brightness, due to the fact that it blocks a part of the screen back-light [Boev 2009a].

Despite these quality differences, both types of auto-stereoscopic displays raise a series of common problems. Among the most important are the alignment of the extra layer of the screen, which must be very precise, the need for displays with very stable pixels (LCDs and plasma screens preferred over CRT displays), or the limitations of the viewing angle [Meesters 2003].

Various methods of overcoming the viewing position limitations have been proposed. The displays can be fitted with *head-tracking* devices that constantly determine the precise position of the viewer and change the barrier arrangement accordingly. Or, the barriers or the lenticular elements can be displayed in such a way that multiple views of the same images could be perceived from different positions in front of the display. The *multi-view display* solution has though its particular disadvantages, since such an approach would mean a decrease in resolution for each of the views. This decrease in resolution is determined by fact that the total number of pixels must now be divided in more pairs of corresponding images, one pair for each view, as Figure 3.9 demonstrates (notations with the same meaning as in the two previous schemas).

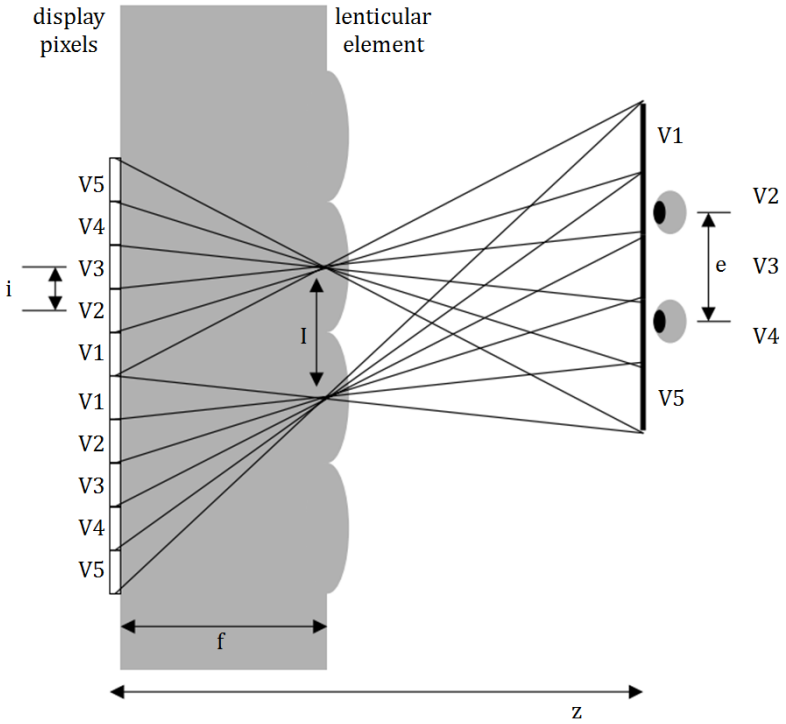


Figure 3.9: An example of a lenticular system with five viewing windows [Holliman 2006].

Due to their particular conditions of visualization that often restrict the number of observers to one, the auto-stereoscopic displays are mostly used for mobile devices. They are also sometimes employed for digital signage applications in shopping malls, museums, or movie theater lobbies, where the high price and narrow sweet spot are less critical [Borel 2013].

3.6 3D data rendering

Since there is a strong interlink between the stereoscopic capture of a certain content and the stereoscopic rendering of that same content, special attention needs to be paid to *correctly adjust the rendering parameters* in function of the recorded 3D data characteristics.

In function of the configuration of the dual-cameras at capture, a given point can be perceived by the viewer either in front of or behind the stereoscopic screen and its depth is directly dependent on its *screen disparity* (also called *parallax*). The perception of a point in depth in function of its positive or negative parallax is explained by Figure 3.10, in which D is the distance from the viewer to the display, d is the metric disparity on the display, e is the inter-pupillary distance, and z is the perceived depth of that point.

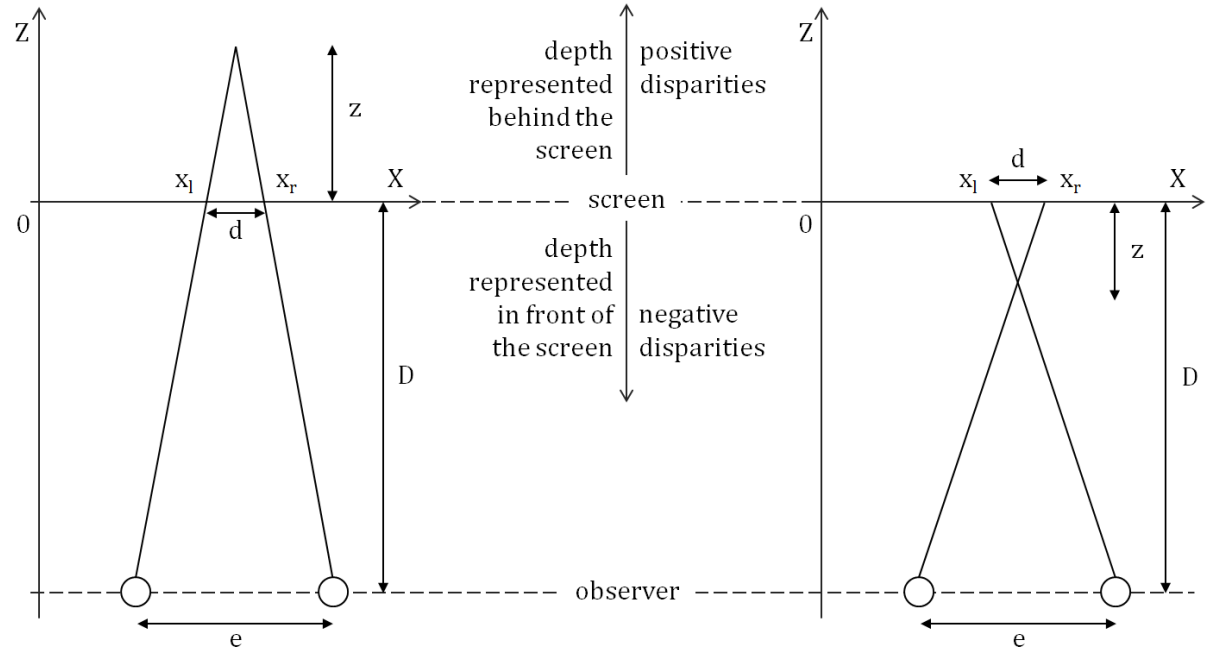


Figure 3.10: The perceived depth of a point on screen.

The *perceived depth*, also called *apparent depth*, can be thus calculated relative to the screen with the following formula [Holliman 2006, Gunnewiek 2010], easily deduced from the geometrical relations in Figure 3.10:

$$z = \frac{d \cdot D}{e - d} \tag{3.1}$$

and using the same notations as for the indicated figure.

Given a certain screen disparity and a fixed distance between the observer and the screen, the smaller the inter-pupillary distance, the more depth is perceived for the same screen disparities, *i.e.* an observer with a small inter-pupillary distance (an important case to take into account is that of children) can perceive larger depths than an observer with a large inter-pupillary

distance. Thus, the Equation 3.1 should be closely observed during rendering in order to ensure that the depth perceived does not exceed the *limits of comfortable viewing* for specific viewers by forcing a too large deviation of their vision mechanisms when watching a 3D screen, relative to the mechanisms used for the everyday natural vision.

In order to better illustrate these constraints, Figure 3.11 represents the perceived depth of a point in function of its screen disparity and illustrates in red the disparity equal to the interocular e. This graphical representation shows best that the disparities on screen should never be larger than this value e, due to the incapacity of the HVS of doing vergence movements that would deviate the directions of sight outward from their parallel configuration.

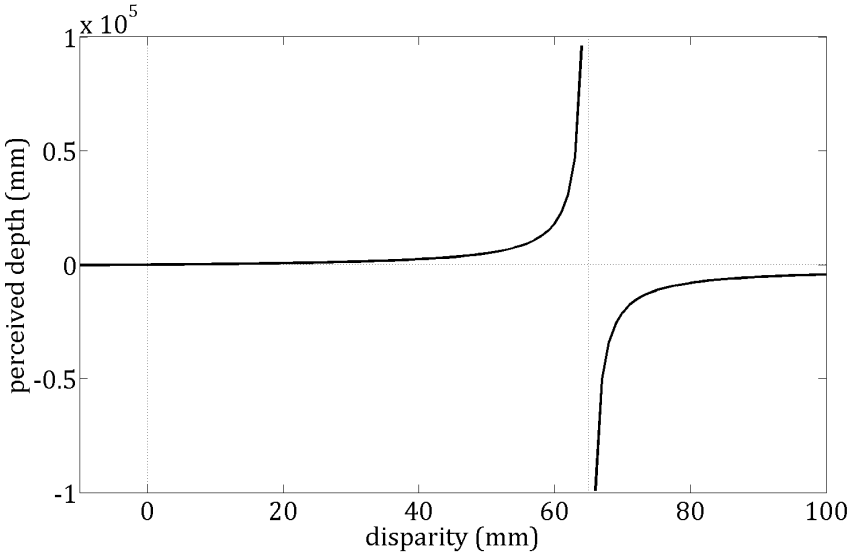


Figure 3.11: The perceived depth of a point on screen in function of the screen disparities.

Rough limits for correct apparent depth amounts in front of or behind the display were determined in literature for a set of viewing distances [Lambooij 2007]. The values were computed based on the fact that the disparity limits and the depth of focus limits almost coincide when measured in distances. The disparity of one degree was considered in this case as the maximum disparity for which depth can be comfortably perceived and the distances obtained are those in Table 3.1.

Table 3.1: Perceived depth limits for a set of viewing distances [Lambooij 2007].

Viewing distance (mm)	Near (mm)	Far (mm)
500	440	580
1000	780	1400
2000	1300	4500
3000	1600	17000

It has been though illustrated in Chapter 2 that the characteristics of the HVS are very complex and vary with a series of parameters. Thus, these given distances are just approximations obtained for the considered maximum disparity and may of course not be suitable for particular

situations.

Besides the focus on the limits of the represented depths, attention must be paid to the *represented proportions* as well. A simple observation related to a correct rendering is explained by the following equation:

$$\frac{f}{w_c} \approx \frac{D}{w_s} \tag{3.2}$$

suggesting that the ratio between the focal length f of the camera and the width of the sensor w_c should be approximately equal to the ratio between the viewing distance D and the width of the display w_s. This is equivalent to saying that the field of view of the scene during capture should be the same as the field of view during rendering [Gunnewiek 2010] and that the proportions in the three directions must generally be preserved between capture and rendering.

A good way of understanding the proportions variation in function of the rendering parameters is to observe that the apparent depth changes with the relative magnitude of the disparities compared to the observer interocular distance. This is illustrated in Figure 3.12 [Starosta 2013] by the different shapes that a perfect sphere can take in function of the different screen sizes that also impose different viewing distances. This schema perfectly shows that the stereoscopic content created for a target screen is not compatible to be displayed on screens of too different sizes.

The explanation on why the depths of the represented sphere vary so much can be given by adapting Figure 3.11 to the three different visualization conditions in Figure 3.12. The graphical representations obtained, reproduced in Figure 3.13, show that the larger screens, imposing larger relative disparities and larger visualization distances, generate perceived depths that vary exponentially in function of the disparities, opposed to only linearly for smaller screens, smaller relative disparities, and smaller visualization distances.

Even when maintaining the display size constant, similar distortions occur when the visualization distance changes. With larger observer-screen distances, larger depths are perceived. Figure 3.14 illustrates this aspect, underlining the importance of a correct visualization distance, adapted to the content displayed.

The presentation of these possible distortions shows clearly the strong influence of the rendering parameters on what an observer perceives on a stereoscopic screen and also how easily the 3D experience can be altered by improper configurations.

Therefore, the design of each stereoscopic system must be nothing but an attentive selection of parameter values for each of the steps in the production-visualization chain, while always keeping in mind the overall picture and the interconnection it imposes.

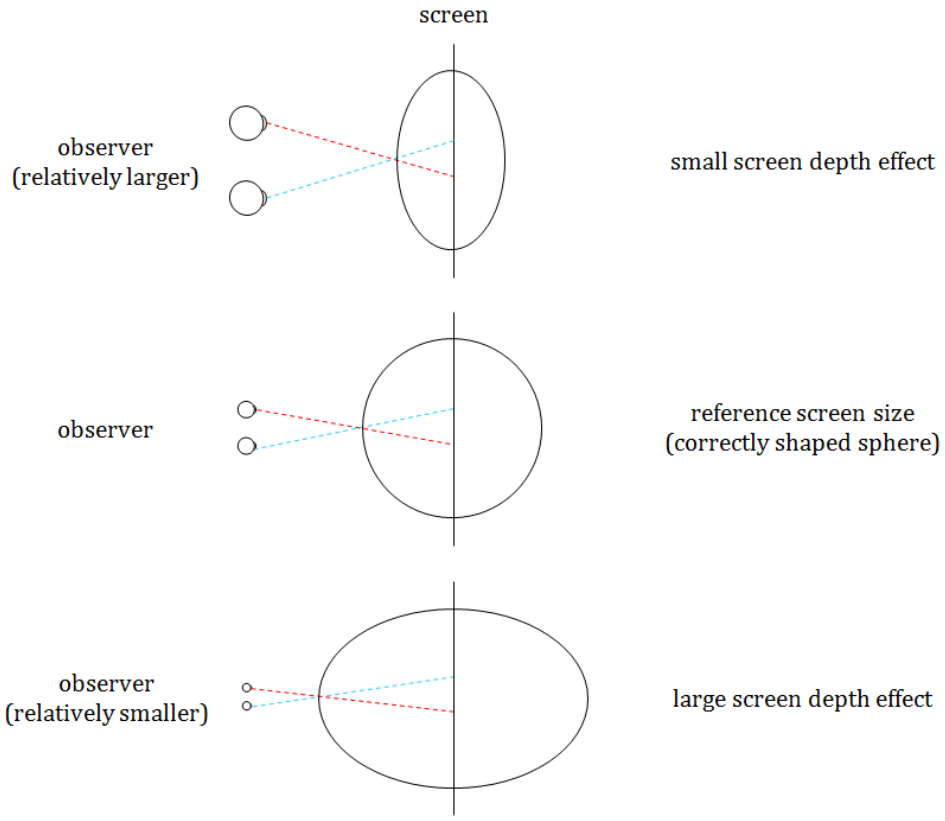


Figure 3.12: The proportion variations in function of the relative magnitude of the disparities on screen compared to the interocular of the observer; adapted from [Starosta 2013].

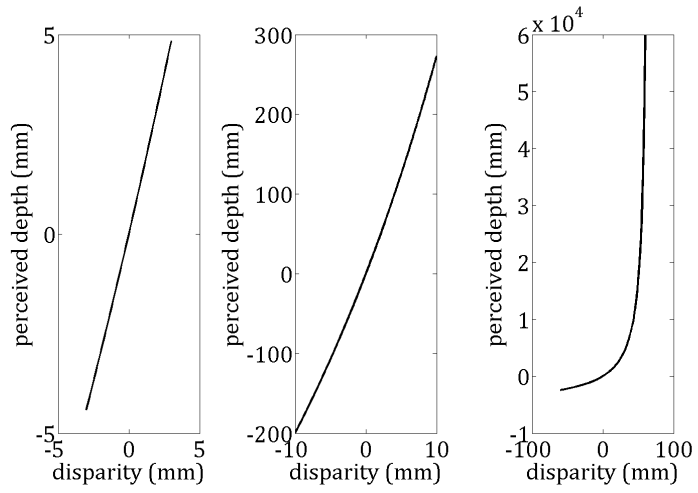


Figure 3.13: The variations of perceived depth in function of the screen disparities, when the relative screen width changes, which imposes changes in visualization distance, and which is equivalent to changes in the relative disparities; representations computed for visualization distances of 10 cm, 150 cm, and 500 cm, from left to right.

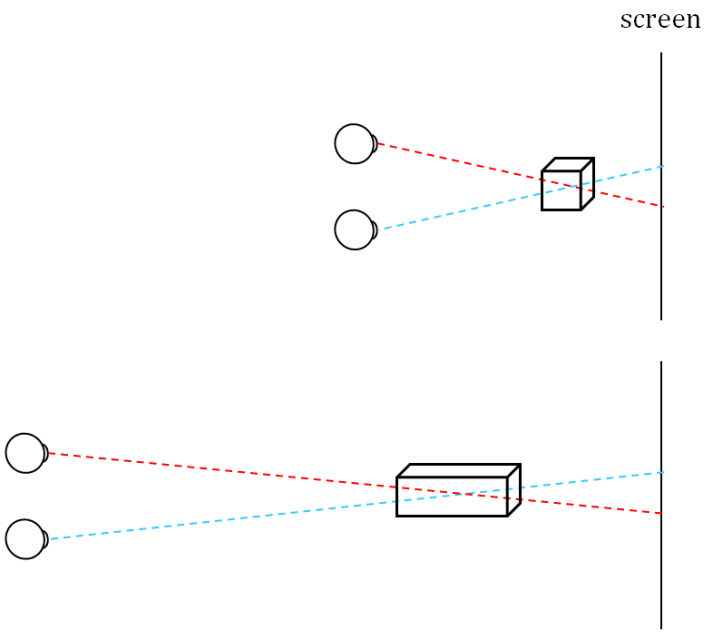


Figure 3.14: The variations in the perceived depth in function of the viewing distance; adapted from [Mendiburu 2009].

Conclusions

The preceding sections were a guide through *the main components of a stereoscopic system* and through the tight interconnections that exist among them.

The explanations included served later in the construction of a stereoscopic system to be used in our studies on the stereoscopic image quality, in Chapter 6.

Chapter 4

The Assessment of the Stereoscopic Quality

Introduction

In order for the various stereoscopic systems presented in the previous chapter to be largely adopted by the public, the *quality* of the images they produce and display must be excellent. But what does *image quality* actually mean? Moreover, what does *stereoscopic image quality* mean? How does the public judge the *visual quality*? And how can this be *measured*?

The present chapter is trying to offer clues for better understanding these issues. However, it is to be stressed that there are no universal answers to such complex questions, but only solutions to particular well-specified problems, as it will be shown in the following lines.

4.1 Image quality

Defining a notion such as *image quality* is not straightforward. The concept is rich and can change its meaning in function of contexts.

In an article concentrated on modeling the image quality, Engeldrum has described this notion as "an outcome of many complex processes that may involve, among other things, software algorithms, chemistry, physics, and the psychology of human judgment" [Engeldrum 2004a]. This definition gives an excellent insight on how image quality is a combination of both measurable physical properties of the images and of more-difficult-to-measure human factors (like human preferences or individual HVS properties that influence how the image properties are perceived).

Therefore, leaving apart the applications where the images are exclusively to be processed by computers (which is not the focus of this work), since the final user of the imaging systems is a person, the human factors always play a very important role in what image quality represents.

In the same article cited previously, Engeldrum sets as well an important delimitation between two large categories of contexts, *i.e.* what he calls the *fitness for purpose* context and the *beauty contest* context. In the first case, the image quality concept comprises the suitability of a given image to a specific application and limits its quality value to that application. In the second case, image quality is a notion intrinsic to a given image, independent of application or utilization case. In any discussion related to image quality, it should be precisely stated which of the two contexts is targeted, since the context changes fundamentally the meaning of the word *quality*.

In the current thesis, we will consider image quality as specified by Engeldrum: "Image quality is the integrated perception of the overall degree of excellence of an image" [Engeldrum 2004a].

4.2 Stereoscopic image quality

Even more complex, the notion of *stereoscopic image quality* (or *quality of stereoscopic images*) complicates things to a greater extent, since the *added value* brought by the *perception of depth* comes into play.

The concept of *added value* can be regarded as the sum of perceptual extra effects created by the stereoscopic combination of the two 2D components that make up the stereoscopic content. These perceptual effects vary from one person to another and are difficult to be represented or precisely measured.

Moreover, the so-called added value of 3D does not always have a positive impact. Even if the sensation of depth usually has a favorable influence on the observer, it seems that the added value of stereoscopic images is only valid when they do not contain impairments or when the level of visible artifacts is low [Tam 1998, Kaptein 2008, Strohmeier 2010b].

Considering the fact that the differences between the perception of 2D images and that of 3D images are given mainly by the so-called added value of the latter, the notion of stereoscopic image quality must encompass thus, on one hand, aspects that relate strictly to the 2D quality of the images that form the stereoscopic pair, but, on the other hand, also aspects that are specific to the human stereoscopic perception of depth.

A generally accepted definition of the *stereoscopic image quality* does not exist yet, to our knowledge. However, numerous concepts have been proposed to refer to the impressions an observer might experience while watching stereoscopic images.

For example, after observing that the term *image quality* "may not be the most appropriate to capture the evaluative processes associated with experiencing 3D images" (fact proved, among other studies, in [Seuntiens 2006b] and [Kaptein 2008]), two new concepts were tested: *viewing experience* and *naturalness*. Experiments proved that both notions were adapted to include the added value of 3D images [Seuntiens 2005, Kaptein 2008], and later a 3D Quality Model was even proposed with *naturalness* as main evaluation concept [Lambooij 2011].

Also, the notion of *presence* (or its equivalent *experience of 'being there'*) was proposed as "a useful criterion to assess the global experiential impact of immersive and perceptually realistic display and communication media in general, and 3D displays in particular" and experimental results showed that the concept was suitable for evaluating stereoscopic images [IJsselsteijn 2002a].

Despite the various vocabulary variations, we consider that Engeldrum's definition on *image quality* can also be adapted to profile the *stereoscopic image quality* in the sense that *the stereoscopic image quality is the integrated perception of the overall degree of excellence of a stereoscopic image*.

As a consequence, in the rest of this text, the *stereoscopic image quality* will always refer to the appreciation of a stereoscopic image in its entirety, and thus avoid the ambiguity that might come from the utilization of other concepts.

4.3 Assessing the stereoscopic image quality

Given these facts and definitions, evaluating the quality of stereoscopic images means first understanding the way human observers perceive the stereoscopic information and judge it as of different levels of *quality*, and then finding the best methods of reproducing this human assessment by standardized means.

Therefore, in the evaluation of the stereoscopic quality, besides the analysis of the 2D com-

ponents, the mechanisms of *depth perception* need to be included as well.

Moreover, whenever speaking of visual quality, numerous visual *impairments* (or *artifacts*) can occur and their effect must be taken into account. The impairments can have an impact on the *visual comfort* or *visual fatigue* of the observer [Shibata 2011, Tam 2011], these being two other concepts of great importance as well in the context of stereoscopic image quality assessment.

These notions will be detailed in the following sections in order to lay the foundations of how the assessment of stereoscopic image quality could be approached. The way these notions can influence the overall impression on quality can be observed later, in Sections 4.7 and 4.8.

In order to bring more precision on this topic, it is to be mentioned here that two distinct approaches exist in the literature regarding the notion of *quality* in the stereoscopic 3D context. One can study the *stereoscopic image quality*, as just described, but one can also study the so-called *quality of experience (QoE)*, which is a more ample notion that takes into account all the elements related strictly to the stereoscopic image quality, but also a multitude of other aspects related to the exterior conditions of visualization (like the ambient lighting or the luminance of the screen) [Chen 2012a]. In the present manuscript, the focus will be entirely on the *stereoscopic image quality* and the studies presented will consider the exterior visualization factors as constant, in order to be able to concentrate exclusively on the quality variations at the image level.

4.4 The depth perceived

The *depth perceived* by an observer when watching a stereoscopic image on a display can be influenced by multiple factors. It can depend directly on the parameters used for capture and rendering (like camera-base distance, screen disparity, viewing distance), on the inter-pupillary distance of the viewer, or even on their pupil diameter [Lambooij 2007]. More aspects on the rendered depth that explain these dependencies on physical parameters have been illustrated in Chapter 2, while the mechanisms that generate depth in stereoscopic systems have been presented in Chapter 3.

However, in a more complicated fashion, experimenters tried to study the dependencies, on one hand, between the depth perceived and the quality of the two 2D components of the stereoscopic data and, on the other hand, between the depth perceived and the so-called image quality of the stereoscopic images. The results so far are often contradictory.

A significant decrease of perceived depth with increasing blur was noted for example in an experiment on still stereoscopic images containing depth and blur variations [Kaptein 2008]. However, other experiments showed that no effect on JPEG coding was found on perceived depth [Seuntiens 2006b], supporting previous similar results stating that the greater depth perceived in the tested stereo sequences did not depend on the visibility of coding artifacts [Tam 1998].

By similar experimental approaches, it was shown that the added value of depth was not always or consistently taken into account when judging image quality [Seuntiens 2006b]. But other studies proved a high correlation between the evaluated image quality and the perception of depth [Aflaki 2010a].

We consider that these controversies come from the fact that the notions used in the evaluations on the stereoscopic image quality are often ambiguous for the naïve observer and that *image quality* for example is interpreted differently from one test to another in the situation of viewing stereoscopic images, in the absence of a standard definition.

What is most important though in the context of our discussion is that it has been shown that the depth complexity of an image influences the stereoscopic image quality (evaluated as the *quality of the 3D experience* in the respective study), if we consider the complexity as determined by the number of objects, depth levels, and motion elements in the image or video [Leon 2008]. Also, in an approach to defining a model of stereoscopic image quality assessment, the relationship between the depth perceived and the *overall 3D quality* was shown to be estimated by a Gaussian model [Yasakethu 2009]. These are key statistical arguments of the importance of including the effect of the *depth perception* in the study of the stereoscopic image quality.

4.5 Impairments in the stereoscopic content

An *artifact* or *impairment* is a visual or perceptual defect introduced during capture, post-processing, compression, transmission, rendering, or display [McCarthy 2010].

The types of impairments that can appear in stereoscopic images are numerous and varied. Most of them are a direct consequence of the presence of artifacts in the 2D components that make up the stereoscopic data, but there are also impairments that are specific to the binocular perception of that data. However, all the impairments may have a considerable impact on the viewer's visualization comfort [Shibata 2011, Tam 2011].

Besides the categorization in function of the source generating them, the stereoscopic artifacts can also be classified in function of the trait that determines the way they are perceived by the human brain: color, structure, motion, and binocular combination [Boev 2008, Mendiburu 2009, Cancellaro 2010, Hanhart 2013]. Another classical categorization is in dependence with the moment in the production chain at which they appear. From this point of view, there are artifacts that are generated during content creation, during coding, during transmission (channel-related artifacts), or artifacts produced during rendering [Boev 2008, Cancellaro 2010].

In the following we present a concentrated enumeration of some of the most important and frequent impairments that are typical for stereoscopic images, as they were presented in a series of comprehensive works on the subject [Woods 1993, Boev 2008, Boev 2009b, McCarthy 2010, Hanhart 2013]. These artifacts are imperative to be taken into consideration when analyzing the effect of stereoscopic content on users [Strohmeier 2010a].

In our presentation, we follow the production chain of stereoscopic images and we illustrate the various impairments in function of the moment when they appear.

Table 4.1 is a useful synthesis of the artifacts to be presented that helps better situating each one of them in function of the classification groups. The following text is in tight relation with the content of this table.

Table 4.1: The classification of the stereoscopic artifacts; adapted from [Boev 2008] and [Boev 2009b].

	Capture	Coding	Transmission	Visualization
Structure	optical blur motion blur noise barrel distortion pincushion effect moustache effect contouring	compression blur block-edge discontinuities mosaic patterns staircase effect ringing aliasing pixelation	data loss and data distortion	flickering aliasing
Color	noise chromatic aberrations vignetting chrominance mismatch	color bleeding		baking long-term use
Binocular	depth plane curvature keystone distortion cardboard effect misalignments unnatural disparities luma mismatch temporal mismatch focal-length mismatch focus mismatch	cardboard effect depth bleeding depth quantization noise depth ringing		crosstalk puppet theater effect gigantism miniaturization pseudoscopy crosstalk

4.5.1 Artifacts produced during capture

The artifacts produced during capture can be a consequence of the properties of each of the two cameras that make up the stereoscopic system, of their joint configuration, or of both in the same time.

At this phase, conventional 2D impairments, like blur, noise, or geometrical distortions are frequent. These 2D artifacts are due either to the optical properties of each of the two camera lenses or to the limitations of each of the two sensors.

Blur appears as a loss of sharpness and details, and thus behaves in most cases as an artifact, diminishing the perceived quality [Kaptein 2008]. It can be generated in different ways. Unwanted blur can be obtained when the recorded scene is out of focus due to improper camera settings and, in this case, the artifact is actually an *optical blur*. It can also be caused by motion, when the exposure time of the photographic camera is too long or, in general, when there is a considerable relative motion between the camera and the scene, both when recording still images and videos. In this case, the artifact is called *motion blur*.

To get into more details, when watching stereoscopic 3D content, the perception of blurred

images can be often caused by inconsistencies between the natural accommodation-vergence reflex of the observer and the specific correlation between accommodation and vergence which is needed for the correct perception of the depth represented on screen [IJsselsteijn 2002b]. Thus, the accommodation-vergence conflict can be a strong cause of perceiving blurred content for inexperienced viewers.

But sometimes blur can even be an important composition element, since, in correctly recorded content, the blur derived from proper DOF adjustments can indicate the precise delimitation of the objects of interest, thus directing the gaze and facilitating the focus on what is important.

The different amounts of blur and its distribution in stereoscopic 3D images can also be a source of ambiguity. Blurred areas of similar appearance can represent both the area in front and behind the object in focus and the polarity of the depth percept might not always be obvious [Lambooij 2007].

Noise or *graininess* is a by-product of low-quality sensors. It appears when their sensitivity or precision is not high enough and artifacts at the color or brightness level are produced in the form of undesired information at certain pixels in the image formed on the sensor.

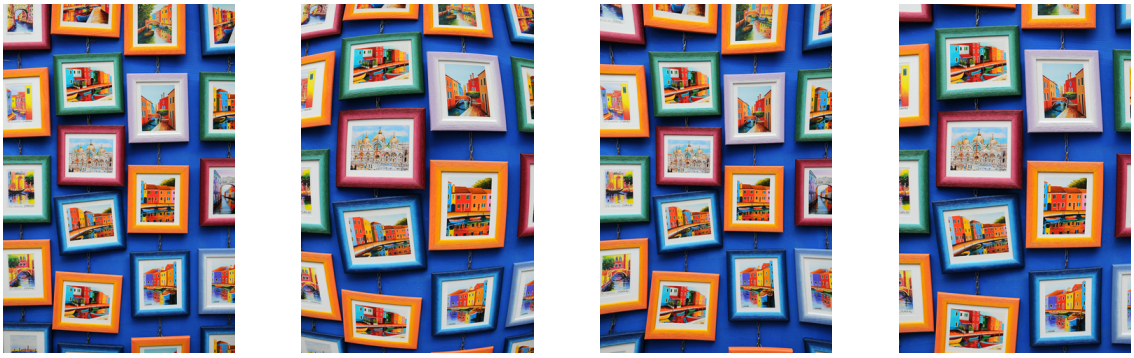
Figure 4.1a shows an original 2D image took by us, while figures 4.1b and 4.1c show the same image to which we added blur and noise, respectively.

(a) Original. (b) Blur. (c) Noise.

Figure 4.1: A 2D image in its original form and with added blur and noise.

If particular types of lenses are used or if the quality of the lenses is low, the *geometrical deformations* of the photographed scene are unavoidable in the recorded pictures. The three most frequent such geometrical distortions are the *barrel distortion* (or *fisheye effect*), the *pincushion distortion* (or *pillow effect*), and the *moustache distortion*, a combination of the previous two. They way each of them deforms a photographed scene can be observed in Figure 4.2.

These distortions can influence the visualization in function of their amplitudes. There always exist differences between the two lenses of a stereoscopic system, in the sense that even if they are close to identical, they are never perfectly identical. Therefore, the different geometrical distortions of each lens lead to mismatches between the two images stored on the sensors when these are binocularly combined. If these mismatches are too large, the disparities thus formed cannot be fused, as it was explained in Section 2.1.5.



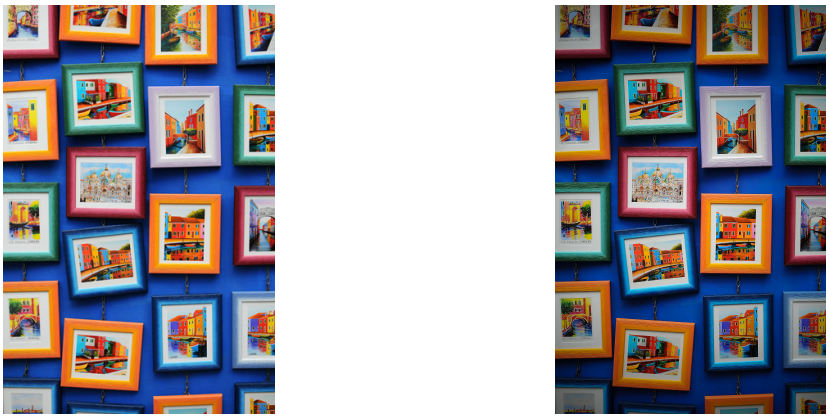
(a) Original. (b) Barrel distortion. (c) Pincushion effect. (d) Moustache effect.

Figure 4.2: The main types of lens distortions.

Many other artifacts can affect the 2D images. For example, *chromatic aberrations*, *contouring*, or *vignetting* have also been found as sources of degradation in 2D images. They also impact the quality of stereoscopic images when these are viewed binocularly. The manifestation of chromatic aberrations is shown by Figure 4.3 and an illustration of vignetting is in Figure 4.4.

(a) No chromatic distortion. (b) Chromatic aberration.

Figure 4.3: A crop of a 2D image, to the left with no chromatic distortions, and to the right with chromatic aberrations.



(a) No chromatic distortion. (b) Vignetting.

Figure 4.4: A crop of a 2D image, to the left with no chromatic distortions, and to the right with chromatic aberrations.

In function of the relative positioning of the cameras during capture, other geometrical aberrations than the ones already mentioned can be produced: the *keystone distortion* and the *depth plane curvature*. These are the most significant geometrical artifacts that appear when using a toed-in dual-camera configuration for recording the data.

The fact that the left and right views of the scene are captured with two toed-in cameras, thus having different orientations, causes the two 2D images of the stereo pair to be projected on the sensors in the shape of two trapezoids, determining the *keystone distortion*, shown in Figure 4.5. The vertical disparities thus generated between the corresponding elements in each projection need to be eliminated through geometrical corrections before superposing the two stereo images correctly. These vertical differences are larger in the corners of the image, proportional to the camera separation, and inversely proportional to the convergence distance and to the focal length [Hanhart 2013].

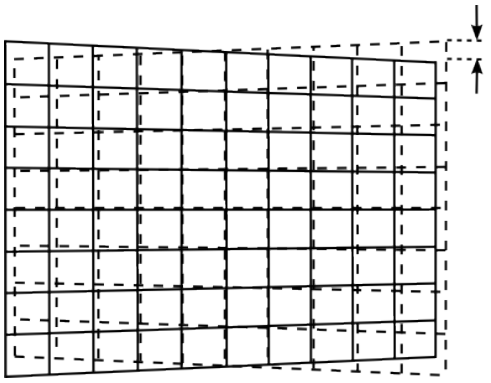


Figure 4.5: Illustration of the keystone distortion [Boev 2008].

The *depth plane curvature* appears in conjunction with the keystone effect and its manifestation is represented by non-uniform ratios between the real depth and the represented depth across the recorded stereoscopic picture. This is again due to the toed-in configuration in which, at capture, as it could be seen in Figure 3.1, the parts of the sensors which are closer to the center of the camera-base distance are slightly farther from the real scene than their opposite extremities, fact that generates stored depths that are deformed as in Figure 4.6.

Both keystone distortion and depth plane curvature can be avoided if using a parallel camera configuration.

The *cardboard effect* occurs when the depths of the objects are not properly represented in the stereoscopic content and the objects appear as parallel slices of cardboard disposed at various depths in space. Figure 4.7 tries to illustrate the manifestation of this artifact on the depth map of a stereoscopic image. The artifact can seem even more pronounced when the observers are very familiar with the exact shape of the objects in reality. The main elements that generate the cardboard distortion are the focal-length of the camera and the object-camera distance, at capture, but this artifact is also related to the rendering parameters, and only a good synchronization between capture and rendering can allow to avoid or diminish it.

Other improper parameter adjustments and improper relative positioning of the cameras at capture can also generate various *misalignments* or *unnatural disparities*. In general, the mismatches of all kinds between the two views of a stereo pair are binocularly perceived as

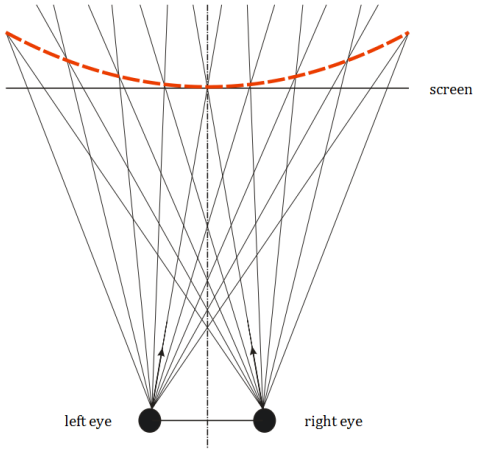
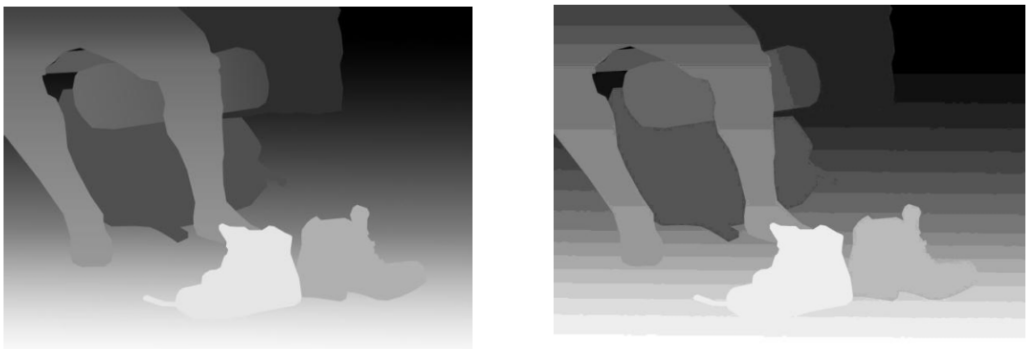


Figure 4.6: Illustration of the depth plane curvature [Boev 2008].



(a) Correct representation. (b) Cardboard effect.

Figure 4.7: The depth distortion produced by the cardboard effect [Boev 2008].

impairments and can cause visual discomfort [Tam 2011]. The inconsistencies can be in terms of color, hue, or saturation (*chrominance mismatch*), in terms of luminance (*luma mismatch*), in terms of synchronization (*temporal mismatch* and, as a consequence of it, *object boundary mismatch*), in terms of capture parameters (*focal-length mismatch*, *focus mismatch*), or in terms of structure (misalignment caused by *rotation* or *vertical disparity*) [McCarthy 2010].

4.5.2 Artifacts produced during coding

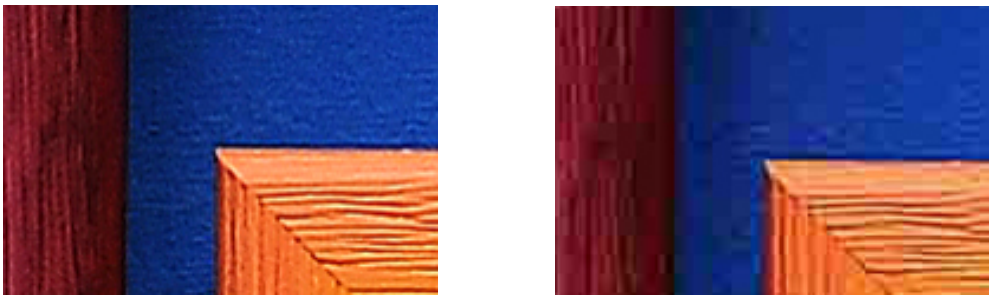
During coding, artifacts can be generated by lossy compression techniques. The choice of the coding format for the captured data determines the compression level and thus the amount of conventional 2D artifacts, like blur or blocking, that appear as a consequence. The coding choice also determines whether disocclusion, particular to specific data formats, will be present as an artifact in the decoded, and then rendered, data.

Blur can thus be as well a consequence of data processing, not only a capture artifact. It can

be generated for example by compression, filtering, or resolution reduction when maintaining the size of the display (or, otherwise, by augmenting the size of the display when maintaining the resolution).

Despite the fact that blur generally degrades visual content, in the case of stereoscopic 3D data, it can be acceptable if using asymmetric coding. This method allows making use of the reduced size of an image that is blurred due to compression by combining it with a very sharp one.

Very numerous, the *blocking artifacts* are a direct consequence of applying block-based compression methods that use the DCT[1]. The most frequent visual effect generated in the case of a consistent compression of this type is that of perceiving *block-edge discontinuities*, which are due to differences in the pixel intensities obtained on the neighboring borders in the processing of each elementary block of pixels. Also, *color bleeding*, *compression blur*, *mosaic patterns*, *pixelation*, *ringing*, or the *staircase artifact* can be consequences of block-based coding. Figure 4.8 shows the blocking artifacts produced after a strong JPEG compression.



(a) Original image. (b) Strong JPEG compression.

Figure 4.8: The blocking artifacts generated by a strong JPEG compression.

This type of artifacts starts to be perceived when the masking capacity of the HVS is outbalanced by the strength of the impairment [Boev 2008]. Since human vision is more sensitive to changes in the low frequency content, the blocking artifacts, like many other similar distortions, are more easily perceived in regions of uniform visual content.

When the depth is coded separately, if its representation is not accurate, or if the content is highly compressed, various depth-related impairments can also occur in the image, like the *depth bleeding* or the *depth quantization noise* [Boev 2008, Cancellaro 2010].

The *cardboard effect* can be generated as well during the coding phase by the coarse quantization of the disparity or of the depth values, in the same case when data formats that store the depth information separately are used.

Also in the situation where the stereoscopic image is coded as a 2D image and an associated depth map, the phenomenon of *disocclusion* appears when rendering the left and the right views of an object that is occluded by another object in the photographed scene. Since the original binocular information is missing, the rendered left and right views will be incomplete in the areas that were originally *occluded* as viewed from the front, from the position the 2D image and the depth map were recorded.

[1] discrete cosine transform

4.5.3 Artifacts produced during transmission

The impairments produced during the transmission phase can be grouped in two large classes: *data loss* and *data distortion*. The artifacts due to transmission errors are sparse, highly variant in terms of occurrence, duration, and intensity, and at very high compression rates may be masked by compression impairments. In general, the channel transmissions always suffer from noise introduction, depending on the technology used and the surrounding environment [Boev 2008].

The data loss and data distortion artifacts may impact both the structure and the color of the images transmitted, therefore they also have an impact on the binocularly perceived data.

4.5.4 Artifacts produced during visualization

There are numerous artifacts that can appear during the rendering of stereoscopic images. They are mostly consequences of the previous stages in the processing chain, that become visible in function of the visualization context. They can also be due to an improper configuration of the rendering parameters, that is not coherent with the capture conditions. Or, the visualization artifacts can be a direct result of the type of display used.

The *visual size distortions*, artifacts related to the capture configuration, are produced if the angular retinal size of a displayed object and its perceived distance do not co-vary as in real world conditions [Meesters 2003].

It is the case of the *puppet theater effect*, a frequent manifestation when using toed-in dual-camera configurations. This is probably the best known size distortion effect and it appears as a disproportionate representation of the sizes of people in particular (or of familiar objects in general) relative to the sizes in the scene, in the sense that background objects do not appear as small as expected, thus foreground objects appear proportionately smaller [Devernay 2010]. It is a consequence of inconsistencies between the interpretation of the binocular and perspective depth cues displayed on screen and the prior knowledge of the viewer about the natural size and depth of the objects in the scene (the more familiar an object is, the more susceptible it is to puppet theater effect) [Boev 2008].

Particular types of visual size distortions, usually intentionally generated for artistic effects, are *gigantism* and *miniaturization*. These distortions are obtained when using a camera-base distance that is different from the average human inter-pupillary distance, considered of 6.5 cm. When the baseline is larger, the perceived sizes of objects are smaller than in reality and, when the baseline is smaller, they are perceived larger [Devernay 2010].

Among the artifacts related to the mismatch between capture and rendering, the *cardboard effect* is the most frequent and it manifests as explained in Section 4.5.1.

Also, in the particular case where the left and right views are accidentally switched at rendering, the depth perception is reversed in the form of *false depth* [Holliman 2006], in the sense that near objects are perceived far and far objects are perceived near. This effect is called *pseudoscopy*, *reverse 3D*, or *pseudo 3D*. Since the resulting stereoscopic image usually makes no sense, the conflicting depth cues telling the brain that the existence of such an image is

impossible make it annoying to watch.

When the left and right views in a stereoscopic pair are not correctly or precisely separated during rendering or, otherwise said, when there is a "leakage of information into the wrong eye" [Woods 2011], the phenomenon of *crosstalk* occurs, as in Figure 4.9. It appears as ghost, shadow, or double contours [Meesters 2003], or as blurring [Lambooij 2007].

The contrast of the display, but also the content of the displayed scene, along with the recording and rendering parameters, such as camera baseline, disparity, or screen parallax, all influence directly the appearance of the crosstalk effect [Boev 2008, Xing 2010a]. The viewer position in front of the screen is of great significance too. An incorrect head position when watching polarized content is a considerable source of crosstalk. This can be avoided though by using circular polarization [Meesters 2003].

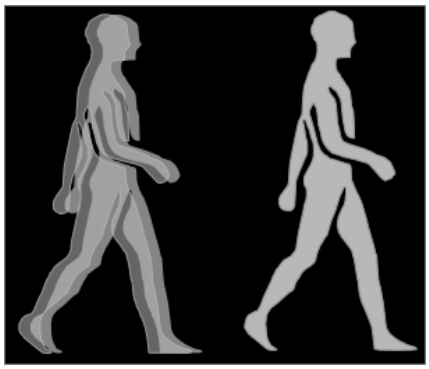


Figure 4.9: Example of the crosstalk effect.

Thus, on different types of displays, the sources of crosstalk are different. In the case of spatial multiplexing, crosstalk can occur, for example, if the observer is at an incorrect viewing distance when the chromatic aberration for lenticular displays is too significant or when there is too much diffraction for parallax barrier displays. Solutions for reducing crosstalk in these cases are either limiting the chromatic aberration or the diffraction, or positioning the observer at the optimal viewing distance. In the case of temporal multiplexing displays, crosstalk occurs mostly if the display persistence is too long. A remedy for this problem is to limit display persistence to very brief durations. Alternatively, one could attempt to measure crosstalk and subtract it from the image of each eye via custom software or hardware [Patterson 2007].

While *crosstalk* can be measured, being the objective and measurable manifestation of the view interspersing, the subjective effect perceived by the observers is often called *ghosting*.

Crosstalk can be annoying to the viewer even when its effect is very little perceptible. Studies showed that high levels of crosstalk reduce considerably the visual comfort [Kooi 2004], but also that the presence of crosstalk can reduce the amount of depth perceived [Watt 2013], and even that the fusion limits are impacted by it, especially for presentations of longer durations [Yeh 1990].

Overall, crosstalk is considered to be one of the main perceptual factors which contribute to image quality degradation and visual discomfort [Meesters 2004]. Nonetheless, a useful utilization of crosstalk also exists. The effect is intentionally introduced in a small amount in certain auto-stereoscopic displays in order to avoid the forming of artifacts like the picket fence effect

(to be discussed in the following paragraphs) or to minimize the effect of artifacts like image flipping (also to be discussed in the following paragraphs) [Lambooij 2007].

The type of display dictates the visual quality of the rendered images to a large extent. The better the display, the fewer the rendering impairments and the better the perceived quality of the represented pictures. Moreover, the strength with which the rendering artifacts are perceived is influenced, besides the type of the display, also by the structure of the visual data displayed, and by the viewing conditions, as the artifacts presented next illustrate.

A frequent, and often unavoidable, rendering problem is represented by the *spatial aliasing* and *color aliasing* (or *false colors*) phenomena. That is, resolution limitations (or previous undersampling during the coding phase) may lead to imprecision in the rendering of high frequency content. Probably one of the best known examples of such manifestations is the *Moiré pattern* [McCarthy 2010].

The *flickering* effect, which manifests as variations in the perceived brightness of an image [Watt 2013], is dependent on the refresh rate of the display. Although it is not a serious problem for the existing 2D high frequency displays, it can be an annoying artifact when watching stereoscopic 3D, in cases where time-multiplexing is used for left-right separation. The effect can be even more disturbing when there is an interference between the display and certain types of illumination sources. For example, neon illumination can generate unpleasant flickering to viewers when they are wearing shutter glasses [Chen 2010].

In strong link with the physical characteristics of the displays, the following parameters are worth enumerating, since they may also influence the quality of the rendered content: the *contrast range*, the *color range*, or the *viewing-angle dependent color representation*. Their limits impact the perception of the displayed content. For instance, for LCDs, the viewing ranges and the contrast ranges are more limited than those of plasma displays [Boev 2008].

Another inconvenience worth to be mentioned is the fact that certain properties of the displays change with the time of use: the *backlight diminishes*, *baking patterns* may be formed, or *dead pixels* may occur [IJsselsteijn 2002b], all these representing visible artifacts on the images that are displayed.

The *picket fence effect* and, in general, the *banding artifacts* are impairments specific to auto-stereoscopic displays. Users perceive an annoying banding of the images, which is determined by the perception of the black barriers that make up the separation between the display pixels that are part of different views (the description of such displays was illustrated in Section 3.5.5).

Image flipping is also specific to auto-stereoscopic displays, but to the particular case of multi-view displays. Along with the banding artifacts, it is perceived when the observer moves his head laterally in front of the display and is a consequence of parallax discretization [Hanhart 2013]. In the case of image flipping, the sensation is of unnatural switch between the adjacent views of a multi-view auto-stereoscopic display, as compared to the smooth natural HVS motion parallax.

The unnatural perception of the proportions of an object when moving the head laterally in front of a stereoscopic display that only allows one fixed viewing position is referred to as *shear distortion*. The disturbance that it causes to the viewer comes in contrast to the natural parallax to which the viewer is used from real life.

Another very annoying particular visual artifact occurs when there are objects of crossed disparity in the recorded scene and they touch the left or the right border of the screen. This is disturbing because a part of those objects is visible only in one view and occluded by the screen frame in the other view. The inconsistency between the depth cues that place that object in front of the screen and the occlusion of that object by the frame, which suggests that the object is represented behind the screen, can thus be a strong source of visual discomfort. This annoyance is reduced for larger displays, case in which the frame is not any more in the field of accurate perception of the observer. The phenomenon is mentioned either as *edge violation*, *window violation*, or *breaking the proscenium rule* [Mendiburu 2009, Devernay 2010].

As it can be seen, the artifacts that have an impact on the visual quality of stereoscopic images are numerous and can appear in a multitude of situations. Some of them can be easily quantified with objective metrics for any image given (like it is the case of blur for example), but others are difficult to measure algorithmically without having a reference image or without owning comprehensive information on the system parameters or on the photographed scene (as it might happen for the geometrical distortions for example).

It will be interesting to see in the following sections how their combined influence on the perceived stereoscopic quality can be approached by the stereoscopic image quality assessment methods.

4.6 Visual fatigue and visual discomfort

Visual fatigue, also referred to as *asthenopia* (literally *eye without strength*) and often associated with stereoscopic displays [Okada 2006], is an overall decrease in the performance of the HVS and is generally directly dependent on the demand exercised on the oculomotor system [Lambooij 2007]. The viewer may not always be conscious of the problem, but its effects can be measured objectively (by assessing for example the changes in accommodation response, the changes of the pupil diameter, or the eye movement characteristics [Tam 2011]). These effects can sometimes manifest on long term as perturbations of the HVS.

Visual discomfort, described in literature as the main health issue in the context of stereoscopic displays, is the conscious and subjective perception of visual fatigue. Generally, perceived visual discomfort determined via subjective measurements is expected to provide an estimation of the objectively measurable visual fatigue [Lambooij 2007].

Comprehensive reviews on the visual comfort and on the visual fatigue have been published in the past [Lambooij 2007, Shibata 2011, Tam 2011], and the following text reproduces some of the key elements they presented on the subject.

The physical symptoms indicating visual fatigue are numerous, from extra-ocular ones (headaches, tiredness, nausea), to eye irritation of various types, or to visual manifestations, like double vision, blurred vision, slowness of focus change, reduced sensitivity to spatial contrast, reduced visual acuity and speed of perception, or reduced power of accommodation and convergence.

Although each of these symptoms can be a strong indicator of visual fatigue, the final verdict can be set only in the presence of multiple symptoms simultaneously. Certain changes in the

oculomotor system can be nothing but a healthy characteristic of our biological system adapting to altered visual environments, and only the changes that are also followed by negative effects on the general comfort of the viewers can be considered premises of visual fatigue.

The factors that cause visual fatigue when watching stereoscopic 3D content can be of different types and can be related to each of the phases of the stereoscopic production-rendering chain.

Some may depend strictly on elements of structure in the 3D content, like high amounts of motion, content composition that breaks the proscenium rule or that includes conflicting cues in general, or rapid changes in accommodation and convergence.The vergence-accommodation conflict can cause, apart visual fatigue, also an increase in the time taken to achieve stereoscopic fusion or a decrease of stereoacuity [Watt 2013].

Other factors influencing visual comfort may be related to the way the content was captured and stored, and a few examples are: content that breaks the horizontal disparity limits or content with vertical disparity or other types of misalignment, camera-base distance improperly set when capturing data, or the presence of keystone distortion or of depth-plane curvature.

A significant influence on the comfort of visualization of the observer have all the various conventional or stereoscopic artifacts that can be present. As an example, blur, especially unnatural blur, is a serious cause of fatigue for the visual system. Excessive blur can make the user feel visual discomfort, but the lack of blur can also have a discomforting effect, since it confuses the observer regarding the zone that he should focus his gaze on, generating an unnatural oculomotor and neuronal effort. This is the case in most stereoscopic scenes, in which the entire stereoscopic image is displayed sharply, because different viewers may concentrate on different parts of the image.

The improper rendering of the stereoscopic 3D content can determine symptoms of visual fatigue as well. When the viewing distance is not respected or when the display technology is of compromised quality, causing flickering or crosstalk, disposing of sub-optimal gaze angles, or presenting other disturbing effects, users may feel discomfort.

The visual comfort can be deteriorated as well by the physical predispositions of each viewer. People with visual anomalies or with binocular visual anomalies, like, for example, convergence insufficiency, accommodation dysfunctions, or uncorrected refractive error, are more expected to feel visual discomfort when watching imperfect stereoscopic 3D images.

The viewing context may generate visual discomfort also. If the observer needs to use special eyeglasses, discomfort might come from simply wearing them. As well, if the observer views the content with an inappropriate head orientation, the perceived distortions of the stereoscopic content decrease their visual comfort.

And, not the least important, the visual-vestibular conflicts or the perceptual and cognitive inconsistencies that might be generated in function of the various stereoscopic contents can also have an influence on the visual discomfort.

All these stereoscopic artifacts that play important roles in the generation of visual discomfort have been discussed in the previous section of this manuscript, Section 4.5.

With the purpose of assessing the visual discomfort and the visual fatigue induced by stereo-

scopic displays and better understanding the enumerated factors that influence them, numerous experiments have been conducted. We only give a few examples of findings and their context.

A study on scenes containing variations of JPEG compression and of camera-base distance at capture suggests that the eye strain felt by the observers of stereoscopic 3D content is directly dependent on the impairments caused both by compression and camera-base distance. The results show a medium statistical correlation between the perceived image quality and the eye strain experienced [Seuntiens 2006b].

Also, concerning the symptoms generated by the distribution of the disparities in the scenes, experiments with random dot stereograms were performed to test whether there is a difference between the influence of negative and positive disparities on the visual discomfort. The results of this study showed no overall differences in symptom severity between the two. However, it seems that the negative conflicts are more uncomfortable at far viewing distances and positive conflicts are more uncomfortable at near distances. The same study also showed how optometric measurements can allow one to predict who will be more uncomfortable and in what circumstances. It thus seems that the phoria and the ZCSBV[2] measurements are predictive of eye tiredness and eye strain, since they are related to the discomfort that the individuals experience in the presence of the vergence-accommodation conflicts [Shibata 2011].

In an investigation concentrated on the evaluation of visual comfort and of the sense of presence for stereoscopic still images extracted from stereoscopic HDTV[3] programs, a strong correlation has been found between visual comfort and the range of parallax distribution, but also between the visual comfort and the minimum parallax [Nojiri 2003]. This comes to support the important influence of the displayed disparities on the comfort of viewing.

An extensive study has focused on the effects on visual comfort of various distortions introduced in stereoscopic still images, like 2D artifacts, misalignments of different types between the two views of the stereoscopic pair, or crosstalk. The results showed that in the case of the 2D artifacts, blur and luster (resulting from the reduction of the color depth of one of the two views) had the highest impact on viewing comfort. Other factors found to influence the visual comfort strongly were visual disparity and crosstalk, even when they were present in little amounts. The distortions that only affected the periphery of the stereoscopic images had less impact on the visual discomfort. So was the case for the binocular misalignments and the excessive disparities for the participants with reduced (binocular) vision [Kooi 2004].

To end this enumeration of experimental results with a more general example, a review of several studies ([Yano 2002, Emoto 2005, Häkkinen 2006a, Jin 2007]) concluded that the viewers have reported in time more visual discomfort and fatigue when watching stereoscopic displays than when watching conventional displays [Shibata 2011]. This shows that the notions of visual discomfort and visual fatigue are still not completely mastered in the stereoscopic context, nor is the overall stereoscopic image quality that incorporates them. Research on the subject is still intensively underway for improvements.

Universally valid methods of stereoscopic image quality assessment are therefore needed for the efficient evaluation of the improvements brought in time to the stereoscopic visual comfort. The next section discusses the state-of-the-art of such assessment methods.

[2] zone of clear, single binocular vision
[3] high-definition television

4.7 Stereoscopic image quality assessment methods

The *image quality assessment methods* in general and the *stereoscopic image quality assessment methods* in particular are solutions mainly adopted by the developers of imaging systems in order to test the capacity of attraction of their products on the targeted public. In order for their systems to be successful, the final quality produced must be outstanding. Therefore, repeated quality tests need to be done at the introduction of any modification in the production chain that could impact the final displayed image.

This section discusses the various methods used for assessing image quality in general and how these can be adapted for the specific case of assessing stereoscopic image quality.

4.7.1 Subjective versus objective approaches

To date, the most accurate ways to evaluate the perceived image quality are the *subjective assessment methods*. Broadly speaking, these imply human assessors who express their opinions on the content under test. They are accurate since they offer the most direct path to associating a quality appreciation to an image, from the true human perception perspective. The drawbacks are that these studies are time-consuming and meticulous, and that they need to be repeated each time a parameter changes in the context under study.

The main subjective quality assessment methods are, on one hand, the highly controlled *psychometric scaling* and, on the other hand, the more supple *exploratory studies*. Besides these, there exist particular methods of subjective exploration, used less frequently, like the *user-centered* approach. Also, there are solutions that analyze the effect of stereoscopic images on the observers through the impact on the visual fatigue. All these methods are described later in this section.

As an alternative to the subjective approaches, the researchers are also focused on developing various *objective assessment methods*. These are standardized algorithms that are based on the characteristics of human perception. They try to deduce a quality rating for a given image only by analyzing the physical structure of that image in an automatized manner that does not imply the presence of any human assessors. However, these objective approaches still use subjective assessments as a validation method of the correct reproduction of the human perception mechanisms by the developed algorithm.

The objective assessment methods have the great advantages of being fast and efficient. Since they are standardized solutions, they can also be used across different studies, allowing for the subsequent comparisons of the results that are obtained each time. They also have an important inconvenient – it is extremely difficult, with the state-of-the-art knowledge of the functioning of the human visual perception and judgment, to produce objective assessment methods that are universal, in the sense of Engeldrum's notion of *beauty contest* situation discussed in Section 4.1. The algorithms proposed so far are mostly context-adapted, which makes their use unfortunately limited to precisely defined situations.

4.7.2 Impairment versus quality approaches

Two classes of situations may be distinguished in the context of image assessment, regardless of the type of the evaluation method (subjective or objective). There is the *impairment approach*, which is suitable in situations where reference images exist and where the images under test need to be evaluated only relative to these references, from which they differ due to a series of impairments. This is why this approach is also called *with reference*. There is also the *quality approach*, in which no reference images exist and the purpose is to obtain an absolute quality appreciation. This second approach is also called a *no-reference* approach [Engeldrum 1999].

4.7.3 Psychometric scaling

Since the human perception is at the center of the image quality assessment, *psychometric scaling* (also referred to as *quantitative approach*) provides methods "to make humans be reliable meters" [Engeldrum 2004b].

Ideally, in order for the research procedures to be reproducible across studies and for the results to be easily comparable, uniform procedures must be followed by all the investigators. Thus, more precisely, psychometric scaling gathers a series of practices that make up a rigorous framework for the evaluations of visual quality using humans. This evaluation framework specifies how the subject selection has to be done, how the stimuli under test need to be prepared, which are the appropriate experimental methods to use, how the experiment plan needs to be built, or how the results of such experiments need to be statistically processed.

In the context of the 2D image quality evaluation, such a precise framework existed for a long time, in the form of ITU[4] Recommendations, among which the most recent and up-to-date is the ITU-R BT.500-13 Recommendation [ITU 2012b]. However, the adaptation of such a framework to the assessment of stereoscopic content was not straightforward. Numerous experiments have been implemented with individual adaptations for a long time, leading to a much too large diversity in the experiment methods that did not allow universal terms of comparison among the studies performed around the world. This happened until the recent proposition by the Video Quality Experts Group (VQEG) 3DTV project [Video Quality Experts Group (VQEG) 3DTV Project 2013] and the ITU-T Study Group 9 [ITU-T SG9: Broadband cable and TV 2013] of a new ITU recommendation at the end of 2012, which is specially adapted to the assessment of 3DTV systems [ITU 2012a] and which will surely facilitate the exchange regarding the stereoscopic image quality results. Detailed information on the principles of psychometric scaling can also be found in Engeldrum's review on the subject [Engeldrum 2001].

Presented next is the general experimentation framework used in most subjective stereoscopic image quality assessments, along with common practices in the field.

[4]International Telecommunication Union

4.7.3.1 Experimental environment

First, the type of images or the type of system under study determines the choice of the *experimental environment*. Two such environments are proposed in the ITU-R BT.500-13 Recommendation – the laboratory environment and the home environment – along with the precise physical properties suggested for the display, like contrast or luminosity, or for the experimental space, like colors or illumination. There are however no such specifications proposed for the studies on stereoscopic images, since most researchers are expected to use current consumer level 3DTV displays, and the characteristics of these vary a lot across manufacturers [ITU 2012a]. The best practices are thus to mention each time the exact experimental configuration in the publications that discuss subjective results.

4.7.3.2 Participants

The choice of the *participants* for the subjective testing sessions must be effectuated with care, since the participants must be a representative sample of the targeted population in order for the results to reflect correctly the quality preferences of that population. They need therefore to be screened before the experiment with various visual standardized tests. Also, the participants usually need to be *naïve* relative to the stimuli to be presented, in the sense that they "should not be, or have been, directly involved, *i.e.*, enough to acquire specific and detailed knowledge, in the development of the system under study" [ITU 2012b]. Statistical data on the participants should be collected during every experiment and summarized alongside the results of the experiment.

The *number of participants* needed in a subjective test is suggested to be of minimum 10 in a detailed review on psychometric scaling [Engeldrum 2001], of minimum 15 in the recommendation for 2D testing [ITU 2012b], and of minimum 30 in the recommendation for 3D testing [ITU 2012a], but successful quantitative explorations have been reported for numbers of observers between 4 and 50 [Engeldrum 2001]. As a superior limit of necessary assessors, the number of 30 participants was given as an example [Engeldrum 2001], but also 20, with the argument that a correct selection of a limited number of participants can be statistically generalized to other subjects having the same characteristics [Hanhart 2013]. Despite these recommendations, it is generally recognized that the number of observers is to be chosen in function of the specific objectives of each study, knowing that an increasing precision of the final estimated results comes with an increasing number of participants.

4.7.3.3 Stimuli

The selection of the *stimuli sample* is considered to be one of the most difficult parts of the experiment preparation, since care needs to be taken of numerous aspects. The factor(s) under study need(s) to vary sufficiently in the experiment stimuli set in order for the differences to be perceptible. These variations brought to the stimuli set represent what is formally called the *test conditions* of the experiment. Also, critical or extreme test conditions need to be included in order to allow the participant to set an anchor for his judgments. Moreover, a single visual stimulus with variations is usually not sufficient. The same test conditions should be applied to several stimuli of assorted spatial configurations, called *test images*, in order for the scene dependence of the judgments to be annulled.

There are two possible configurations proposed in the ITU-R BT.500-13 Recommendation for the design of the stimuli set and they are schematized by the equations:

$$stimuli\ set = test\ images = test\ conditions = F_1 + F_2 + ...F_n \tag{4.1}$$

and

$$stimuli\ set = test\ images \times test\ conditions = test\ images \times (F_1 \times F_2 \times ... \times F_n). \tag{4.2}$$

Either each test image represents one level of one factor only (Equation 4.1), or each test image represents one level of every factor examined but, across images, each level of every factor occurs with every level of all other factors (Equation 4.2) [ITU 2012b].

The great advantage of the second configuration is that it allows the detection of interactions among factors.

4.7.3.4 Experimental methods

A wide variety of subjective assessment methods have been used so far in the investigations on conventional image quality or stereoscopic image quality. These methods define the way the set of stimuli are displayed for each observer, for which duration of time, and in what order. They also give indications on the way the assessment should be made by the participants and on the ratings scale to be used.

In function of the way stimuli are presented, there are two large classes of methods. The first is the *single stimulus* category, which implies that only one stimulus is displayed on screen at a given time and that the participant needs to attribute a rating to it. The second category is the *double stimulus* one, where two stimuli are displayed simultaneously at a given time. The two stimuli are either a reference image and a distorted one, or any two random images from the stimuli set. The task of the participant, in the case where the reference image is always present, can be to rate the distorted image in function of the first one. Or, when two random images are displayed, their task can be to select the preferred image from each couple of images displayed. These are however only basic guidelines, the experimental method chosen can deviate from these principles in function of the objectives of each particular study.

A distinction among the possible assessment methods can be made by the way the ratings are cast by the observers. There exist *categorical scaling* methods or *continuous scaling* methods, in function of type of the rating scale made available to the assessor. But the word "continuous" can also be used in the context of *continuous evaluation* or *continuous rating*, referring to the experiments on video sequences, where the participant can effectuate the assessment during the whole duration of one video sequence, opposed to only once at the end. Since only focusing on still images, in the current work we will not refer to the latter type of methods.

Regarding the specific terminology, the smallest structure that composes an experiment is called *assessment trial* and represents the evaluation by an observer of a single image or of a single group of images simultaneously with a unique *opinion score*. Also, the assessment trials are usually grouped in *test sessions*, which contain the totality or a part of the assessment trials that make up the experiment. The maximum suggested duration for a test session is of half an

hour, therefore the experiment needs to be segmented in more then one test sessions only in the case where the total experiment duration is longer than that. A general rule is that a random order should be used for the arrangement of the assessment trials in a test session.

At the end of the experiment, the individual opinion scores can be concentrated into *mean opinion scores (MOS)*. However, if reference images are included in the test session(s), the *difference opinion scores* between the scores of the distorted images and the scores of their corresponding references can be calculated, then the *diffrence mean opinion scores (DMOS)* obtained.

Table 4.2 concentrates a varied sample of existing subjective assessment methods, indicating for each one the situation for which the method would be suitable and also the document that describes it. The table is interesting especially for illustrating that the range of existing methods is large and that the choice of the appropriate assessment solution should always be done attentively according to the context of the study.

Table 4.2: Summary of subjective assessment methods; adapted from the ITU-R BT.500-13 Recommendation [ITU 2012b].

Assessment problem	Method used
measure the quality of systems relative to a reference	the double stimulus continuous quality scale (DSCQS) method
measure the robustness of systems (*i.e.* failure characteristic)	the double stimulus impairment scale (DSIS) method
quantify the quality of systems (when no reference is available)	the ratio scaling method or categorical scaling
compare the quality of alternative systems (when no reference is available)	the method of direct comparison, the ratio scaling method, or categorical scaling
establish the point at which an impairment becomes visible	threshold estimation with the forced choice method or the method of adjustment
determine whether systems are perceived to differ	the forced choice method
measure the quality of stereoscopic image coding	the double stimulus continuous quality scale (DSCQS) method
measure the fidelity between two impaired video sequences	the simultaneous double stimulus for continuous evaluation (SDSCE) method
compare different error resilience tools	the simultaneous double stimulus for continuous evaluation (SDSCE) method

Besides the assessment methods summed up in this table, in the ITU-R BT.500-13 Recommendation, four alternative methods or method groups are illustrated. These are the *single stimulus (SS)* methods, the *stimulus comparison (SC)* methods, the *single stimulus continuous quality evaluation (SSCQE)* method, and the *simultaneous double stimulus for continuous evaluation (SDSCE)* method.

Since among all the methods enumerated so far, only four have been presented in the ITU recommendation for tests on stereoscopic content [ITU 2012a] and among the four, only three are suitable for the evaluation of still stereoscopic images, we only concentrate on these three methods in our present state-of-the-art presentation.

The single stimulus (SS) method. This procedure is characterized by the presentation of a unique stimulus to an observer who has to judge it by selecting a rating.

The set of stimuli should be prepared as described in Section 4.7.3.3, choosing from one of the two configurations proposed, in the case where the influence of more than one factor on the perceived quality needs to be assessed.

The structure of a trial assessment proposed for the evaluation of stereoscopic images is made up of three displays. The first one is a pre-exposure mid-grey field that may contain a fixation target of less than 3 seconds in duration. The second one is the stimulus to assess, presented for 5 to 10 seconds, in function of the particular experimental situation. And the third display is a post-exposure mid-grey field during which the vote should be cast, shown for a longer duration, in order to make sure the participant has the time to register his rating. Figure 4.10 gives the schema of this trial assessment configuration.

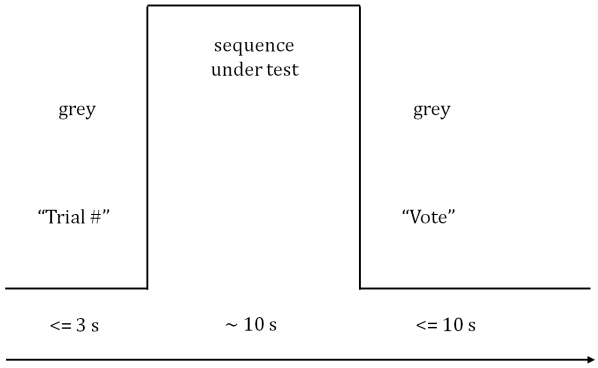
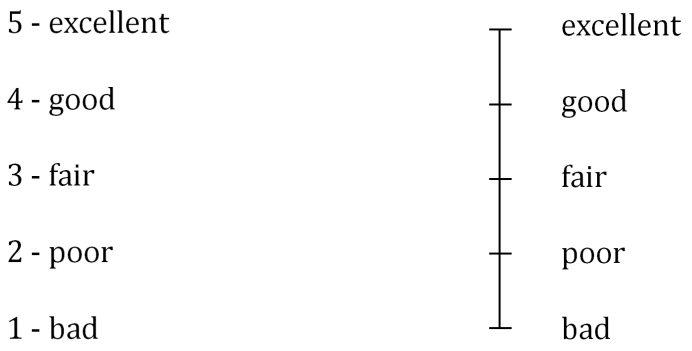


Figure 4.10: The structure of a trial assessment for the SS method[ITU 2012b].

The trial assessments can be presented either only once in the test session, or three times, forcing the test session to be segmented in three presentations, each of them including all the images to be assessed only once. Tripling the number of appearances of each trial assessment allows to stabilize the observer's opinion with the first presentation, while considering for data processing the mean of the last two presentations.

In order to express their quality judgments, in this method the participants are presented with the five-rating labeled scale illustrated in Figure 4.11, that can be used either in a discrete (Figure 4.11a) or in a continuous (Figure 4.11b) manner. Other choices of scale are possible

also, such as numerical categorical scales or non-categorical scales, but the first solution is the preferred one, being the only one maintained in the recommendation adapted to stereoscopy. The reason for this might be the fact that, by using a scale presented both in an adjectival and in a numeric form, on one hand, the perceived intensities to judge can be named and, on the other hand, an equal spacing between consecutive words on the scale is ensured.



(a) Discrete scale. (b) Continuous scale.

Figure 4.11: The labeled five-rating ITU scale for the subjective assessment of (stereoscopic) image quality [ITU 2012b].

In the cases where the illustrated scale is used in its discrete form, it becomes a categorical scale and the SS method can also take the name of *adjectival categorical judgment* method or *absolute category rating (ACR)* method.

For the processing of the experiment results, either the MOS or the DMOS values may be computed.

The double stimulus continuous quality scale (DSCQS) method. In the DSCQS trial assessment structure, the participants are presented alternatively with a reference image (A) and an impaired image (B). If there is only one participant watching the images at a time (variant I of the protocol), the participant can switch freely between the A and B stimuli for the number of times desired. However, this method can be implemented also with multiple observers watching the images in the same time (variant II). In that case, the A and B stimuli are alternated automatically for a limited number of times. Successive stimuli are always separated by a mid-gray display of 3 seconds, and the stimulus display duration can vary from 3-4 seconds for still images to 10 seconds for video sequences. The number of repetitions of the A-B alternation for variant II is suggested to be of five for still images (with voting during the last two) and of only two for video sequences (with voting during the second repetition). The succession of displays during a trial assessment is presented in figure 4.12.

Both the reference and the impaired images need to be assessed in the DSCQS procedure and the scale recommended is the five-rating scale proposed for the SS method, in its continuous form, as in Figure 4.11b.

The processing of the results is done by obtaining first the difference opinion scores between the reference and impaired stimuli, and then by computing the DMOS.

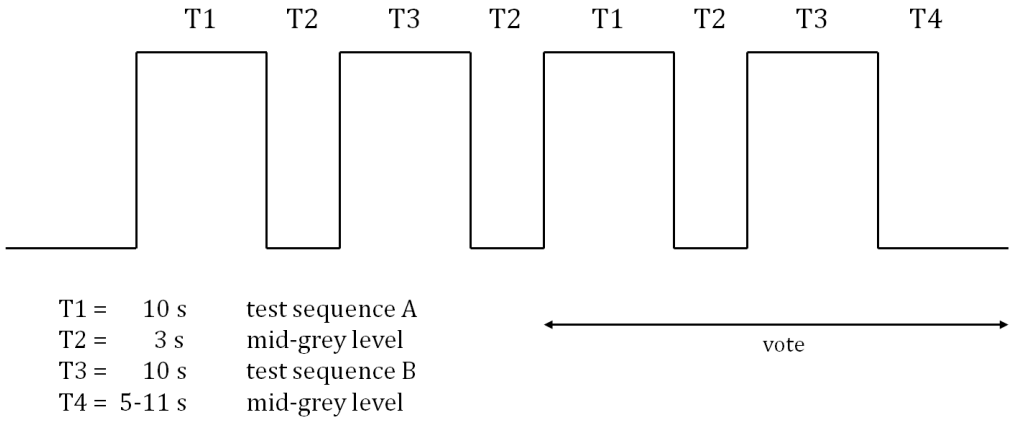


Figure 4.12: The structure of a trial assessment for the DSCQS method [ITU 2012b].

The stimulus comparison (SC) method. This protocol, also referred to as the *paired comparison (PC)* method, consists of a series of trial assessments during which the participants needs to compare two images displayed simultaneously, preceded and followed by mid-gray displays, exactly as in Figure 4.10, corresponding to the SS method. The two stimuli can be also presented sequentially, with a 3 seconds mid-gray display between them, and this configuration can be seen in Figure 4.13.

The stimuli must be two randomly chosen images of different test conditions. The number of trial assessments needed in one experiment is the one that covers all the combinations of any two such stimuli.

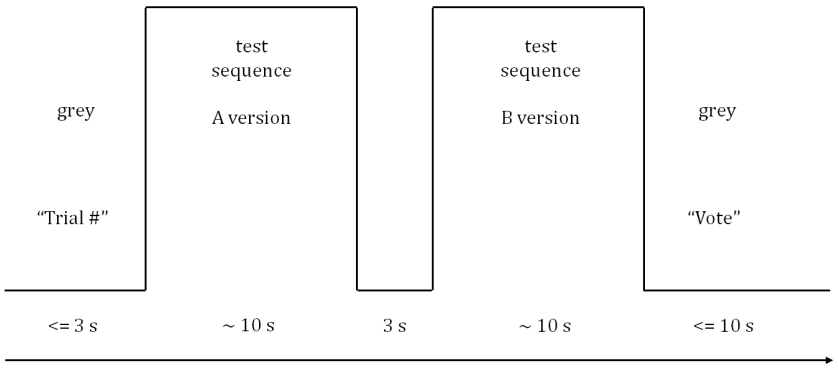


Figure 4.13: The structure of a trial assessment with a sequential presentation of the two stimuli for the SC method [ITU 2012b].

Since the judgments in this protocol are in terms of preference, they can be expressed either using a binary scale, or by giving a graded preference on a scale like the one in Figure 4.14.

-3 – much worse

-2 – worse

-1 – slightly worse

0 – the same

1 – slightly better

2 – better

3 – much better

Figure 4.14: The labeled ITU graded scale for the subjective assessment of (stereoscopic) image quality with the SC method [ITU 2012b].

4.7.3.5 Statistical data processing

Of great importance among the steps to be performed during an experimental study is the statistical processing of the gathered data. This step also is to be done according to the specificity of the investigation context, of the participants, or of the form of the recorded values. However some guidelines are offered in the ITU recommendations relative to the computation of the MOS or of the corresponding confidence intervals. A method of screening the participants for eliminating the ones with scores deviating from the general tendency is proposed as well [ITU 2012b]. We included these mathematical aspects in Appendix B.3 of this manuscript.

4.7.3.6 Experiment script

The *experiment script* or the *test plan* is a written sequence of procedures, questions, or instructions to be followed by the experiment administrator, that is recommended to be appended to the description of every experimental study [Engeldrum 2001]. It can also include details on the environment configuration or on the software used, or any other information that would help to characterize the specificity of the experimental approach. Such a detailed plan can have multiple benefits, among which the facility to test and modify it in order to fine-tune the procedure, the possibility of using it for reproducing exactly the same experimental protocol in time or at different locations, and also the complete documenting of the study. Moreover, it also helps the experimenter not to deviate from the standard procedure, thus avoiding the *moderator bias*.

To date, it is not a usual practice to share the experiment script in the publications describing the results, but doing so would surely contribute to a better documented literature of results and would facilitate the comparison among studies.

4.7.4 Exploratory studies

When modifications are brought to well established systems and the factors to test are precisely known, psychometric scaling is surely adapted to assessing the impact of those new changes on perceived quality.

However, there are cases where the bases of a new framework need to be set and where the impact the new framework might have on human observers is completely unknown. In such cases, preliminary investigations need to be made on the new system, and these are often made in the form of *exploratory studies* or *descriptive studies*.

The exploratory tests are thus studies during which no direct subjective ratings are acquired, but instead unprimed attitudes, feelings, and reactions towards the new type of content are explored [Meesters 2004]. These aim at identifying and understanding what perceptual attributes are relevant to the users in order to determine their quality preferences [Hanhart 2013]. As a consequence, due to their focus on defining or understanding, rather than on quantifying, this type of studies is also referred to as *qualitative*, as opposed to the quantitative psychometric scaling.

The familiarity with the stereoscopic systems has not yet reached a satisfactory level from the point of view of the impact they have on human perception. Therefore, the exploratory approaches have been found well suited to get a deeper understanding of the added value and underlying perceptual constructs of the novel stereoscopic technologies [Meesters 2004], before proceeding to more precise psychometric evaluations.

A series of qualitative tests have been reported up to now in the literature. Due to their exploratory nature, they can be very different one from another, using methods like focus groups, think-aloud (or co-discovery) procedures, or interviews and interpretations of various types [Meesters 2004, Jumisko-Pyykko 2007a].

In a broad review on research methodologies for experiments, the two large categories of *qualitative interviews* and *sensory profiling* were proposed to structure a series of exploratory studies reported so far [Jumisko-Pyykko 2008].

The qualitative interviews have been used up to now either independently or in conjunction with quantitative evaluations. In the second case, they are applied usually after the latter and consist of a series of questions destined to reveal underlying factors of subjective quality [Jumisko-Pyykko 2008]. As opposed to structured interviews, in which the set of questions is fixed and does not allow distraction from it, the semi-structured qualitative interviews have been preferred due to their open structure and to their reported characteristic that the effects of the interviewer is reduced [Jumisko-Pyykko 2007a, Jumisko-Pyykko 2007b]. An example of questions asked in a semi-structured interview on audiovisual quality is given in Figure 4.15.

In the sensory profiling category, tasks like free-sorting and description, or free-choice profiling are included. An interesting example of study that includes free-sorting and description is the *Interpretation Based Quality (IBQ)* approach. During a subjective experimentation on images containing sharpness variations, the participants had the task of classifying the images according to subjective similarities, without being told which was the impairment factor the experiment was focused on. Subsequently, the participants had to describe the reasons for the classification they had effectuated and the type of differences distinguished. By effectuating a correspondence

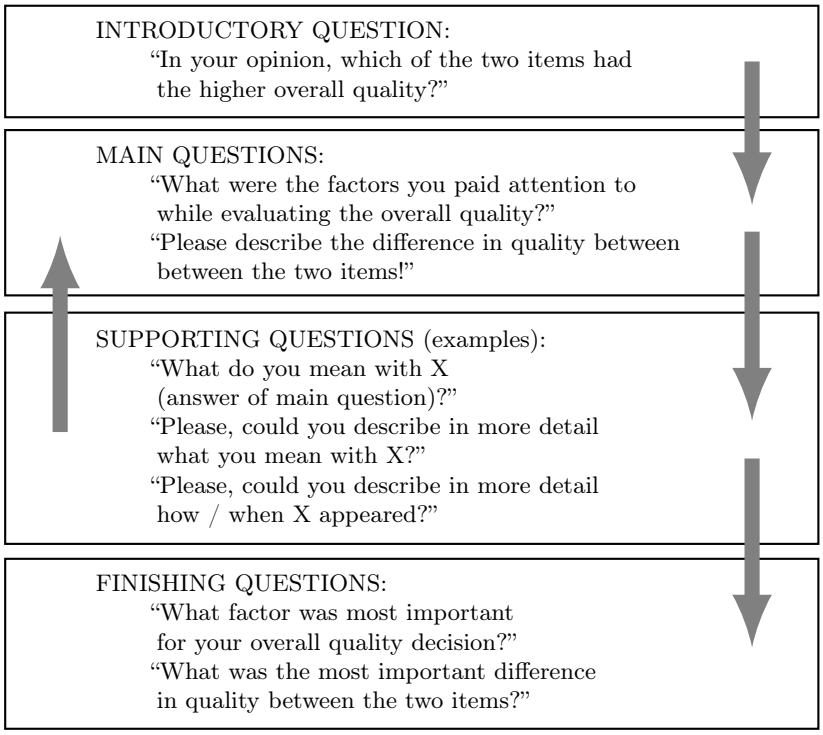


Figure 4.15: Example of a semi-structured interview [Jumisko-Pyykko 2007b].

analysis between these results and the results of a psychometric test evaluating the sharpness of the images under study, the conclusion was that strikingly different attributes had been associated with each sharpness level. These findings illustrate how the IBQ approach allows to study what subjective attributes underlie the decisions that individual observers make when estimating images of different quality [Nyman 2006] and, consequently, how the exploratory studies allow preparing the premises for better psychometric scaling.

4.7.5 User-centered studies

In a large sense, all the image quality assessment approaches that take into account the human perception can be seen as *user-centered* (or *human-centered*) approaches. However, in the context of the exploratory studies, the user-centered methods are particular solutions that relate quality assessment to the different aspects of use, referred to in the literature as the *context of use (CoU)* [Jumisko-Pyykko 2010a]. Consistent such studies have been developed for example for the evaluation of the improvements that can be brought to mobile 3D TV systems, given their utilization in different contexts of use [Gotchev 2008, Jumisko-Pyykko 2010a].

The CoU-based user-centered studies are particularly elaborated and time-consuming, due to the large amount of data that needs to be collected (one set of data per each utilization context considered) and due to the complicated analysis to be effectuated on that data. We do not focus on this type of protocols in this work, however, we show an example of an experiment plan designed in this context in Table 4.3.

Table 4.3: Example of a user-centered experiment plan [Jumisko-Pyykko 2010a].

	MACRO LEVEL context of use	MICRO LEVEL context of use
PLANNING	1. User requirements 2. Characteristics of CoU (form) Analysis of threats of validity	
DATA-COLLECTION Procedure -Pre-test ↓ -Context 1 ↓ -Context 2 ... -Context N ↓ -Post-test	 3. Quality evaluation task 4. Characteristics of CoU (Semi-structured observation, moderator) 5. Experiences of CoU (interview) 6. Workload (NASA-TXL questionnaire) 7. Experiences of CoU (interview) 8. Experiences of CoU (interview)	 9. Situational data-collection (Audio/video analysis)
ANALYSIS	10. Quality evaluation (statistical analysis) 11. Realized characteristics of CoU 12. Workload (statistical analysis) 13. Experiences of CoU (Data-driven analysis) Integration of results	15. Situational data (Data-driven analysis)

4.7.6 Studies centered on visual fatigue

Focusing on the human factors perspective is considered of essential value in the studies on the acceptance of stereoscopic technologies [Meesters 2004]. In this setting, subjective and objective methods exist for quantifying the effect that the stereoscopic content has on the physical symptoms of the viewers [Lambooij 2007].

There are no standard methodologies for the measurement of the visual comfort levels generated by stereoscopic images [Tam 2011], but all the subjective assessment methods presented previously in this chapter can be adapted to its evaluation. Moreover, specific objective measurement methods have been developed in time for the quantification of visual fatigue. The three large groups of objective solutions presented in a review on visual fatigue and visual discomfort are *measurements made with optometric instruments*, *measurements made with clinical optometric diagnoses*, and *brain activity measures*.

In the present work, we focus our attention on the Simulator Sickness Questionnaire (SSQ), a subjective investigation method that allows to estimate the physical symptoms that are likely to deteriorate after the immersion in a simulator-like experience. These physical reactions are thus called *simulator sickness symptoms*. The SSQ was originally developed as a method of evaluating the effects of aviation simulator displays on a series of physical symptoms [Kennedy 1993]. It was however soon used, as-is or adapted, for evaluating human factor aspects related to various other consumer immersive technologies.

The SSQ was used for example in an adapted form in order to illustrate the differences in the human effects determined by the use of a HMD[5] as a *personal display system* (head-tracking disabled), in contrast to the effects determined by a classic display. The authors expected to find that the sensory conflict specific to the personal display system (for head movement, no visual response indicating this movement) would generate stronger symptoms than a classic display. Their results confirmed the hypothesis, with augmented scores in the case of the personal display system for general discomfort, fatigue, headache, nausea, dizziness, and stomach awareness [Howarth 1997].

When analyzing the symptoms after watching a movie on a television, watching the same movie on a HMD, and playing a game on the HMD, the use of the SSQ allowed to conclude that playing the game on a HMD produced the strongest levels of nausea and disorientation. However, surprisingly, more nausea was recorded after watching the movie on the classic television than on the HMD in the considered test conditions [Häkkinen 2002]. In a study where the participants played two video games (a fast-paced car racing game and a simple and stationary car racing game) on two different displays (a head-worn display and a CRT[6] display), the authors observed that simulator sickness symptoms were perceived in all the four cases, showing that a captivating and audio-visually appealing game does not decrease the reported simulator sickness symptom levels. Since the levels of nausea and disorientation were larger for the fast-paced game than for the simple one in a similar way on the two displays, it was also concluded that the main difference in the context of the experiment was between the game types and not between the display types [Häkkinen 2006b]. A third study on video games showed that for the same game played on a CRT display and on a head-worn virtual display with or without stereoscopy, the strongest symptoms were recorded for the stereoscopic mode on the head-worn display, with only small differences of symptoms between the two cases of 2D displays [Häkkinen 2006a].

Also, through a group of five experiments performed on auto-stereoscopic displays, during which the SSQ was administered several times, a slight and mainly short term increase of the simulator sickness symptoms could be observed [Jumisko-Pyykko 2010b]. The findings indicate that, for certain visualization conditions, an adaptation can occur that leads to a reduction in intensity of the symptoms for a longer duration of the immersion.

A certain number of experiments thus focused on the impact of watching stereoscopic video media on various types of screens and showed that the simulator sickness symptoms can be influenced by the type of the display, by the content watched, by the activity performed by the observer, and by the duration of the visual immersion.

The exact form of the Simulator Sickness Questionnaire and the method for processing its results are detailed in Appendix C.

[5] head-mounted display
[6] cathode ray tube

4.7.7 Objective assessment methods

Given the complexity, the long durations, or the planning and environment constraints implied by the subjective assessments presented in the preceding sections, it is certain that a more practical method for measuring stereoscopic image quality would be that of using standardized objective methods that could be applied no matter the context and no matter the particularities of the stereoscopic images under test.

Unfortunately, the existence of such "universal" algorithms seems an utopia and the research results to date are far from accomplishing the conditions enumerated. Further more, even if the approaches proposed so far are numerous and covering a large variety of perspectives on the perceived stereoscopic image quality, the majority of the proposed metrics include the study of only a few of the factors that influence the human perception and do not encompass the entire complexity of influences and inter-correlations between these factors. Moreover, most of these metrics are only suitable for utilization in precisely defined contexts, as also explained in Section 4.7.1.

In the following, a compact review of the various types of metrics proposed so far for the objective evaluation of stereoscopic images is presented. It is structured in function of the approach used, in the sense that certain experimental studies tried to use 2D metrics for the assessment of stereoscopic data, other studies tried to adapt the 2D metrics in original ways in order to make them more suitable to the stereoscopic evaluation, and other studies used new approaches based on the characteristics of the HVS.

The impairment and the quality approaches discussed in Section 4.7.2 are illustrated in this presentation of metrics, in the sense that the metrics that need reference images are usually used to evaluate the presence of distortions, while the others try to asses the quality in its absolute meaning.

4.7.7.1 Using conventional 2D metrics

In time, there have been many attempts of exclusively using classical 2D quality metrics, like PSNR[7], SSIM[8] [Wang 2004], or VQM[9] [Pinson 2004], for the evaluation of stereoscopic content. Each of these metrics has been applied independently on the two 2D images (or 2D videos, in the case of VQM) making up the stereoscopic pair, then the results have been combined, generally by averaging.

On one hand, in the case of stereoscopic videos, some correlations between the results obtained with 2D metrics and the results of subjective assessments on the same content have been found. For example, average VQM was found to predict very well the overall quality for sequences impaired by different coding levels, in comparison with other metrics. Also, average SSIM proved to be a slightly better indicator of perceived depth, relative to the other studied metrics [Yasakethu 2009]. Other research results showed that VQM is a suitable metric for the evaluation of both image quality and depth perception for asymmetrically coded video sequences [Hewage 2008].

[7] peak signal-to-noise ratio
[8] structural similarity index
[9] Video Quality Metric

On the other hand, multiple research results showed that certain 2D metrics are not suitable for analyzing stereoscopic quality when different compression thresholds are used (*e.g.* PSNR applied to stereoscopic still images [Gorley 2008] or to stereoscopic video sequences [Joveluro 2010]). Furthermore, several researchers have even reached the conclusion that the conventional 2D metrics in general cannot be adopted in evaluating the stereoscopic image quality directly, when the stimuli are degraded by blur, compression, or noise [You 2010].

Another experiment tested the use of four 2D quality metrics (SSIM, UQI[10] [Wang 2002], C4 [Carnec 2003], and RRIQA[11] [Wang 2005]) for analyzing the quality of stereoscopic images containing different impairments caused by coding and blur. The metrics were applied separately to the left and right view, then the results were combined in three different manners called the average approach, the main eye approach, and the visual acuity approach. The general conclusion was that no metric performs acceptably for all kinds of distortions, despite the fact that C4 proved good results when assessing JPEG coded images. Also, the last two approaches did not improve the results obtained with the average method [Campisi 2007].

Considering that the best way of studying the human perception of quality would be to take into account the HVS characteristics, it was suggested to use the PQM[12] 2D metric for the assessment of stereoscopic 3D content. This metric measures distortions in luminance and contrast and proved to be in high correlation with the results obtained from subjective experiments, when compressed video sequences were assessed [Joveluro 2010].

A combination of three structural similarity components was also proposed as a quality assessment method for stereoscopic 3D images. It was based on the premise that different types of degradation result in different effects on images. The left and right images were thus decomposed in edge, texture, and smooth region information and, for each type of information, left SSIM and right SSIM were weighted by a general formula that modeled the importance of each analyzed part in the overall perception. The proposed schema showed good consistency with the subjective results in the case of stereoscopic images with impairments caused by blur, white noise, JPEG, and JPEG 2000 compression [Mao 2010].

4.7.7.2 Adapting conventional 2D metrics

A slightly different approach compared to the previous one is that of using the 2D metrics in combination with other evaluation methods or metrics in order to build an assessment solution that is closer to the complex human perception.

A straightforward approach for refining in such a way the assessment of stereoscopic image quality was the attempt of combining a measure of the image quality with a measure of the depth sense of the stereoscopic content. The image quality was computed as the average of the left and right PSNR values and the indicator of depth quality was given by SSA[13], a metric that determines the PSNR between the significant disparity points of the original disparity image and the corresponding disparity points in the impaired disparity image. This method proved good correlation with the results obtained through subjective assessment [Yang 2009].

[10] Universal Quality Index
[11] Reduced Reference Image Quality Assessment
[12] Perceptual Quality Metric
[13] Stereo Sense Assessment

In a similar experimental study, numerous 2D metrics were applied both to the two 2D views and to the disparity information of impaired stereoscopic images. The values that were thus obtained were compared to the values derived from the original unimpaired content and would lead to a measure of the decrease in quality. With this approach, the metrics were applied either in a local manner, in a global manner, or in more complex ways that combine the two. The researchers concluded that the results were promising, but they remarked that further study was needed for a better understanding of the HVS system and for finding an improved way of combining the disparity information and the 2D image quality information extracted from the data in a true stereoscopic metric [You 2010].

In another study with the same purpose of jointly using the assessment of the 2D image quality and the assessment of the quality of the disparity information, two different manners of combining the two types of results were proposed. The first method made the average of the results obtained for each view of the stereoscopic pair (using SSIM or C4), then fused this 2D image quality score with the value representing the disparity distortion, leading to a global measurement of the stereoscopic image quality. With the second method, disparity distortion measurements performed locally were combined separately with the left and right image quality scores and were only afterwards combined by averaging in a final and general result. After correlating the results of the proposed metrics with those obtained through subjective assessments, the main observations were strictly related to the 2D metrics used. C4 had already shown good results even without considering the disparity information and no improvement was present when the score of the disparity distortion was taken into account. In the case of SSIM though, the inclusion of the disparity distortion evaluation brought an improvement, showing that, from a perceptual point of view, besides luminance, contrast, and structure, depth also has a strong influence on quality perception. A more general view on the experiment results led to the conclusion that the depth information can improve quality metrics, but the relation with image naturalness, viewing experience, and presence still has to be investigated in depth, being dependent also on the different stereoscopic display technologies used. Moreover, it was again obvious that "a complete analysis which could bring the definition of a universal and objective quality metric for quality of experience assessment for stereo images and video is still far to come" [Benoit 2008].

4.7.7.3 Other metrics

The SBLC[14] algorithm was proposed as a perceptual quality metric that could be successfully used for predicting a threshold compression level for stereoscopic images. The algorithm takes into account characteristics of human perception, like the fact that the HVS is more sensitive to contrast changes than to changes in luminance or that it has varying sensitivity to areas with varying spatial frequency. Therefore, the metric detects regions of high frequency in the stereoscopic image pairs and measures the changes in contrast and luminance between the original and the impaired content for those specific areas. After subjective experiments, it was concluded that the quality levels detected with SBLC outperformed the precision obtained with PSNR [Gorley 2008].

An approach that went into the depths of neuronal processes, down to cell level, was the one which led to the elaboration of a quality metric based on binocular energy. The proposed

[14]Stereo Band Limited Contrast

algorithm models the response of vision cells responsible with binocular fusion, when different contents are displayed. The behavior of the cells was simulated by applying the CWT[15] on the views of a stereoscopic pair, and then by applying the bandelet transform on the coefficients obtained from CWT. This second step led to results representing the size and orientation of every simple cell, where identical phase and orientation for corresponding cells would mean no disparity. The matching of the complex cells was realized with a binocular energy model which was applied to the original images and to the degraded ones and the final score given by this metric was calculated as the difference between the binocular energy of each pair of images. The performance of the metric was declared encouraging by its authors [Bensalma 2010].

In a study from the perspective of the impact of capture and rendering artifacts on the perceived stereoscopic quality, camera baseline and crosstalk were considered as two factors that could give relevant quality indications. Consequently, an objective metric that measured their effects was proposed. The metric made use of the SSIM measure in order to model the effect of crosstalk as perceived by the HVS and also of a perceptibility threshold for weighting the perceived crosstalk. The overall viewing experience was then regarded as a combination between the SSIM map and the disparity map. The metric proved high correlation with the results of subjective testing on the considered stimuli set, outperforming the use of classical 2D metrics like PSNR and SSIM [Xing 2010a].

A similar approach by the same authors was that of considering scene content, camera baseline, and screen size as the three factors to be exploited for a measure of the visual quality of experience. In this case, a correlation of 87 % with the subjective results was determined for both tested databases, better than the results obtained with classical 2D metrics. Integrating the proposed metric with other distortions in the processing chain was declared as a future perspective [Xing 2010b].

In the form of a no-reference approach, a metric based on 2D and 3D natural scene statistics was proposed. For its implementation, the 2D features must be extracted from the cyclopean image generated from the disparity map and from Gabor filter responses of the stereo pair. Also, the 3D features must be determined from the disparity map. The 2D and 3D features are then fed into a quality estimation module which predicts the perceived 3D quality. The proposed metric significantly outperformed stereoscopic no-reference algorithms and delivered competitive performances relative to high-performance stereoscopic reference algorithms [Chen 2013].

And a last example of an objective metric is that of an algorithm that stresses the distinction between the *information-loss distortions*, like blur or magnification, and the *information-additive distortions*, which modify the inherent information of the original images, like the Gaussian noise or blockiness. This algorithm makes use of the fact that the perceived quality of a stereoscopic pair is dominated by its high-quality component when information-loss distortions are present and does not follow the high-quality component in cases where information-additive distortions are introduced. Therefore, it was proposed to compute the luminance, contrast, and structural similarities between the original and the distorted stereoscopic images, then derive binocular luminance, contrast, and structural values from these, and finally integrate the three into an overall quality index, by using appropriate weights and by taking into account the principles stated. It was declared that the proposed metric can provide consistent and outstanding performances with both symmetric and asymmetric coded stereoscopic images, compared to four other existing methods [Ryu 2012].

[15] complex wavelet transform

Given these several illustrations of metrics used so far in the evaluation of the perceived stereoscopic image quality, the view that the elaboration of a universal objective metric seems an impossible goal is reinforced. Despite the improved performances that the objective methods have reached in time and despite their impressive performances in certain contexts, there is still no solution that is able to perform a correct assessment regardless of the type of the images, of their capture parameters, of their processing, or of their rendering. The human perception of the stereoscopic images remains an infinite universe to be explored.

4.8 Stereoscopic image quality models

What has been presented so far are only small pieces of the very wide field of stereoscopic image quality assessment. The numerous results reported account only for solutions to isolated problems and this is not sufficient. In order to aim to real practical applications, comprehensive models that integrate the complex dimensions of what the stereoscopic image quality represents are necessary.

One can still doubt if the existence of a model that predicts human perception from image characteristics is possible, but in an interesting experimental approach, it has been shown that indeed the visual quality experience can be predicted by using measurable physical and computational level features [Eerola 2008]. This experiment concentrated on print quality, but its generalized conclusions can stand true in any context in which visual quality needs to be assessed.

Thus, one of the current challenges in the context of stereoscopic systems today is to elaborate such a stereoscopic image quality model, by identifying the appropriate perceptual attributes that define stereoscopic image quality and by finding the influence of each one of them on the global percept.

In the present text, we illustrate the few stereoscopic image quality models proposed so far in the literature to our knowledge.

First, Seuntiëns proposed the *3D visual experience* model reproduced in Figure 4.16 as represented by the joint contribution of the underlying attributes image quality, depth, and visual comfort [Seuntiens 2006a]. This same model was studied later by [Chen 2012b].

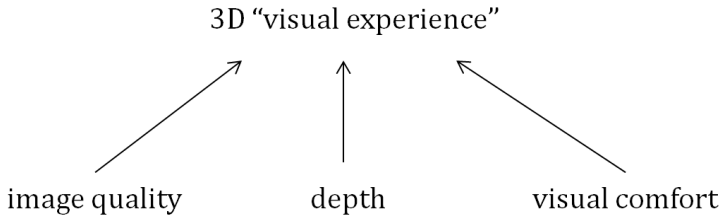


Figure 4.16: The first 3D visual experience model proposed by Seuntiëns.

Following a series of experimental studies by Seuntiëns, this model became more complex, as represented in Figure 4.17. The 3D visual experience was thus defined by two ample concepts,

naturalness and visual comfort, while naturalness was determined to be a combination of image quality and depth variations [Seuntiens 2006a].

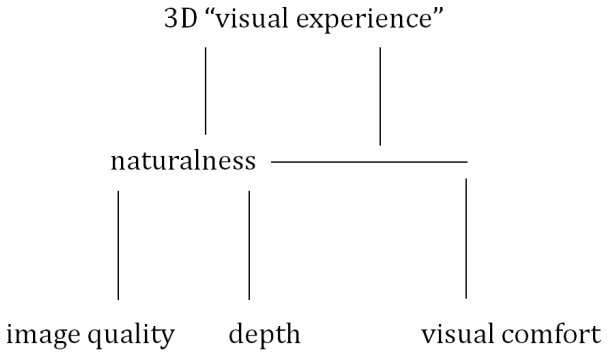


Figure 4.17: The improved 3D visual experience model proposed by Seuntiëns.

A study by Strohmeier *et al.* added precision to Seuntiëns' model, in the sense that, through a sensory profiling approach, it was shown that depth perception only seems to contribute to the added value of the presented material when the level of the artifacts is low. This indicates a hierarchical, rather than equal, dependency of the visual experience on the underlying attributes considered [Strohmeier 2010c].

The model by Seuntiëns was further simplified by Lambooij *et al.* into the *3D quality model* shown in Figure 4.18. The subjective experiments they carried out have shown that the concept of naturalness is more appropriate to evaluate the added value of stereoscopic depth in still images than the concept of viewing experience and also that both concepts follow the variations in image quality and in stereoscopic depth introduced in the stereoscopic stimuli. Moreover, precise weighting parameters could be derived for the influence of both image quality and depth and these are 0.74 for the former and 0.26 for the latter, respectively [Lambooij 2011].

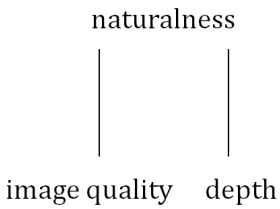


Figure 4.18: The 3D quality model proposed by Lambooij *et al.*.

This very short review illustrates how important steps have been made recently into the understanding of how the various perceptual attributes integrate into the multidimensional concept of stereoscopic image quality. However, the limited number of such proposed solutions leaves this particular issue almost unexplored and declares the challenge open for many numerous similar investigations to come.

4.9 The Image Quality Circle of Engeldrum

The Image Quality Circle (IQC) (Figure 4.19) is a robust framework, or formulation, which organizes the multiplicity of ideas that constitute image quality [Engeldrum 2004b]. It is a framework that relates the technology variables of an imaging system to the customer quality preferences and thus allows modeling the judgment of image quality independent of references [Engeldrum 1999].

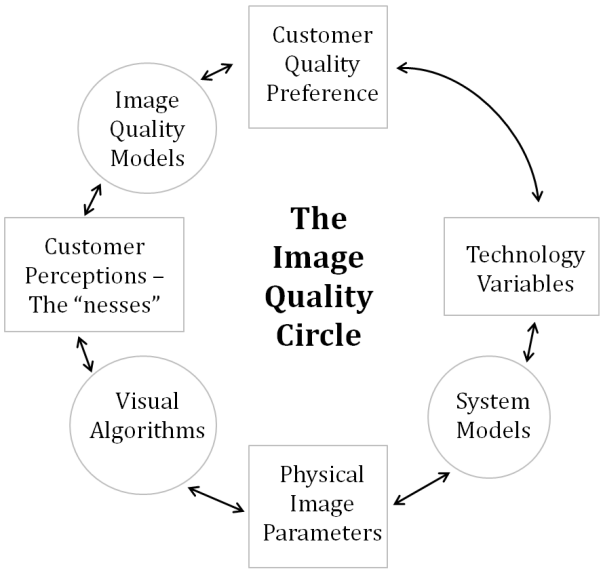


Figure 4.19: Engeldrum's Image Quality Circle

This framework was first presented in the context of print quality [Engeldrum 1989], but its generalized schema is suitable for modeling principles related to quality for a large range of other systems involving visual content, including stereoscopic systems. Several studies have already used this framework for developing theories related to stereoscopic image quality [Seuntiens 2006a, Lambooij 2011].

The following presentation of the IQC theory reproduces a series of concepts and definitions, exactly in the form given by Engeldrum, that offer a very useful and accurate formalization for the field of image quality assessment [Engeldrum 2004b]. We find that the definitions and the theories proposed are an appropriate manner of summing up all the information presented in this chapter and that they compensate successfully the necessity of "tidying up" a field where many isolated studies have been proposed, many new vocabulary terms adopted, many models developed in time, but in which the global view has been missing for a far too long period of time.

Image = *a colorant arranged in a manner to convey* information *to a human observer. Colorant is used in its most general sense. It can be ink, plastic (toner), wax, dye, silver, phosphors, light emitters, and so on.*

Image quality = *the integrated perception of the overall degree of excellence of an image* (in the already discussed *beauty-contest* sense).

Customer image quality rating (the image quality value) = *the judgment a customer renders for a sample image, independent of context, application, and fitness for purpose.*

Technology variables (the things we control) = *the items that are used to describe an imaging product, such as dots per inch or pixels per inch, image sensor megapixels, paper thickness, or the waterfastness of the image materials.*

Customer perceptions (the *nesses*) = *the perceptual attributes, mostly visual, that form the basis of the quality preference or judgment by the customer. A* percept *is a sensation or impression received by the mind through our senses. An* attribute *is a characteristic of the image. So a perceptual attribute, or* ness*, is a characteristic of an image that we sense (see). Most visual perceptual attributes associated with imaging end with the suffix* ness*, so this is the telltale clue. Some examples are sharpness, graininess, colorfulness, lightness, and brightness.*

Physical image parameters = *quantitative and usually obtained by physically measuring the image with instruments or computations on an image file.*

System/image models = *formulas, physical models, algorithms, or computer code that connect the physical image parameters and the technology variables.*

Visual algorithms = *formulas, models, or computer code that connect physical image parameters to customer perceptions; recipes that are used to compute a value of a* ness *from a physical image measurement. An example might be computing sharpness from the gradient of an edge image.*

Image quality models = *link customer perceptions, the* nesses*, with customer image quality ratings. The image quality model inputs are values of* nesses *and the output is the customer image quality value. This is the ultimate destination in the many-to-one mapping process of the image quality circle: a one-number summary description of image quality.*

Along with these concepts and definitions, the following important guidelines specific to the Image Quality Circle framework are worth to be mentioned:

- *Customer image quality ratings are independent of application.*
- *The concept of IQC is based on judgments (which are objective), and not on opinions or preferences (which are subjective).*

- *Customers do not judge image quality or technology variables at all. Instead, observers make a quality judgment, or express a preference for a particular image, based on what they see – the* nesses.
- *Most of the model construction relies entirely on psychometric scaling studies.*

Having placed here these formal premises of the Image Quality Circle of Engeldrum, light has been cast on the logical schema of the relations that connect the technology variables to the customer ratings in any imaging system. This framework also shows how the approach on subjective assessments should be correctly performed and, most importantly, explains the logic behind building image quality models. The part on the system/image model formalism is of less interest in the particular context of this thesis.

Although references to the Image Quality Circle have already been made in the literature on the stereoscopic image quality, no thorough adaptation of the IQC to the stereoscopic context has yet been proposed.

Conclusions

This chapter allowed to sum up the current practices for *the evaluation of the stereoscopic image quality*. It discussed definitions, methods, and research results. It explained the parallel between the *subjective* and the *objective* solutions of evaluation and also the differences between the *impairment* and the *quality* approaches.

The detailed description that it offered on the *stereoscopic image quality assessment methods* allowed us later to make a reasoned choice for the methods to use in our studies, presented in Chapters 8, 9, and 10.

The state of the art presentation of the existing *stereoscopic image quality models* showed that the questions related to them are still not completely answered and justified our orienting of our work in that direction.

This chapter was also the opportunity to present the *Image Quality Circle* of Engeldrum. We were able to adapt this framework to the context of stereoscopic image quality and showed later, in Chapter 5, how the new *Stereoscopic Image Quality Circle* is suitable to place at the right position each of the pieces in the puzzle of the logic behind the human perception of stereoscopic quality.

Part II

Our Objectives and Our Approach

Chapter 5

Our Objectives and Our Approach

Introduction

The work presented in this manuscript stands, on one hand, as an attempt to better understand how the *stereoscopic image quality* is *perceived* and *judged* by human observers and, on the other hand, as a step forward in the *modeling* of this concept.

The main purpose of this chapter is to clearly state these objectives that we set in our work and to introduce the tools that we used in our approach to attain them.

5.1 Our objectives

Given our focus on the human perception and judgment of the *stereoscopic image quality*, we have initially formulated two main objectives for our work and they can be described as follows.

Our first objective was to identify the *main perceptual attributes* that form the basis of the multidimensional concept of *stereoscopic image quality*, from the human perspective. This meant understanding which are the elements on which the human observers base their judgments when asked to assess the complex notion of stereoscopic image quality.

Our second objective was formulated as a direct consequence of the accomplishment of our first goal. Once the base perceptual attributes identified, we had the purpose of finding a suitable *mathematical framework* that would model the *relations* among these *perceptual attributes*, the *physical properties* of the stereoscopic images, and the *overall percept* of stereoscopic image quality.

This could be realized in *three steps*.

First, *the role of each perceptual attribute* identified should be understood and explained *relative to the overall percept* of stereoscopic image quality.

Then, *the perceptual attributes* should be *related to the physical properties* of the stereoscopic images.

And, finally, *a global view*, regrouping the previous two points, should be elaborated, in order to check if the multidimensional concept of *stereoscopic image quality* could be *estimated in function of the physical parameters* that define the structure of the stereoscopic images.

Our defined *work plan* was therefore drew as illustrated in Figure 5.1 and the rest of this manuscript will detail our approach to accomplish the objectives set in it.

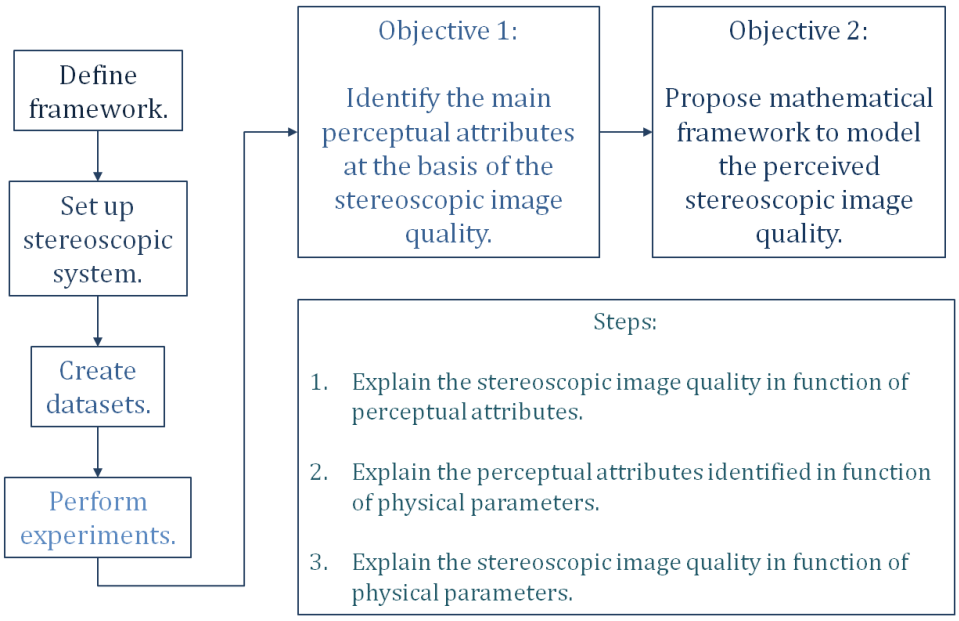


Figure 5.1: Our work plan illustrating our main objectives.

5.2 Our approach

A series of research *tools* have proved their utility for us on the path to finding answers to our formulated objectives.

First, a *framework* was necessary in order to better structure the mechanisms implied by the assessment of stereoscopic images. This framework was adapted by us from the IQC[1] of Engeldrum proposed in the context of classical 2D images.

Second, a complete *stereoscopic system* was needed in our laboratory, that would reproduce all the production chain steps, from capture to visualization.

[1]Image Quality Circle

Third, *datasets* of precise structures had to be elaborated, using the stereoscopic system implemented.

And fourth, an exploration of the human judgment on the stereoscopic image quality and its composing attributes could only be done through the means of *subjective experimental studies*, that would use the datasets previously created.

The following sections introduce our use of each on of these tools to attain our goals.

5.2.1 The framework – the Stereoscopic Image Quality Circle

As a first tool, we considered the IQC of Engeldrum for conventional images, presented in Section 4.9, as an optimal starting point and an appropriate framework to be adapted to the stereoscopic context as well, due to its qualities. The IQC is in the same time a simple, clear, compact, and universally adaptable framework for the image quality assessment, and also an efficient way to structure the complexity of its underlying processes.

As we already argued in Section 4.9, this framework can be successfully adapted to model the processes related to the assessment of any type of visual quality. Therefore, we considered it naturally suitable for modeling the stereoscopic image quality assessment as well.

We consequently proposed an adaptation of the Image Quality Circle framework to the context of stereoscopic images, in the form of the *Stereoscopic Image Quality Circle (SIQC)*.

To this end, we initially show how the elements presented in the state-of-the-art part of this manuscript (Part I) integrate in the SIQC framework proposed. Then, we introduce the adaptation of the definitions referring to the original case of conventional image quality to the particular case of stereoscopic image quality.

Since the logic of the structure behind it does not change, the schema synthesizing the SIQC framework, shown in Figure 5.2, maintains the same form as for the IQC.

5.2.1.1 The components of the SIQC

In Figure 5.2, the *technology variables* refer to the physical characteristics of the stereoscopic systems presented in Chapter 3. They are the capture, processing, or rendering parameters that define the outcome of that system. These parameters can be used as entry data for the *stereoscopic system/image models*, which help schematize how the technology variables of a system determine the physical properties of the stereoscopic images created with that system.

However, these two key notions are of secondary importance in the context of the present work, which is less focused on the technical aspects of the stereoscopic systems, but mostly centered on the mechanisms of human perception of given images with given properties. Of firsthand importance are thus the notions related to the SIQC that we explain next and that we highlight in Figure 5.3, which shows the key elements of the SIQC on which we concentrated during our work.

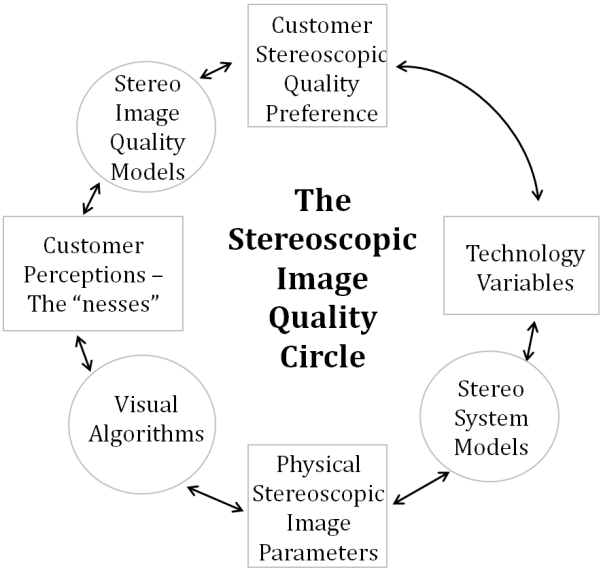


Figure 5.2: The Stereoscopic Image Quality Circle

The objective assessment methods illustrated in Section 4.7.7.1 can be naturally associated to the *visual algorithms* component of the SIQC, still with one deviation from the IQC structure. This deviation stands in the fact that most of the objective assessment methods for stereoscopic images proposed so far have tried to relate the *physical stereoscopic image parameters* (*e.g.* pixel intensity or disparity) directly to the *customer quality ratings*, which are collected with subjective methods like those presented in Section 4.7. In the meanwhile, the *customer perceptions*, tightly related to the mechanisms of human perception presented in Chapter 2, have usually been neglected and the intermediary step represented by the *stereoscopic image quality models* has almost never been taken into account.

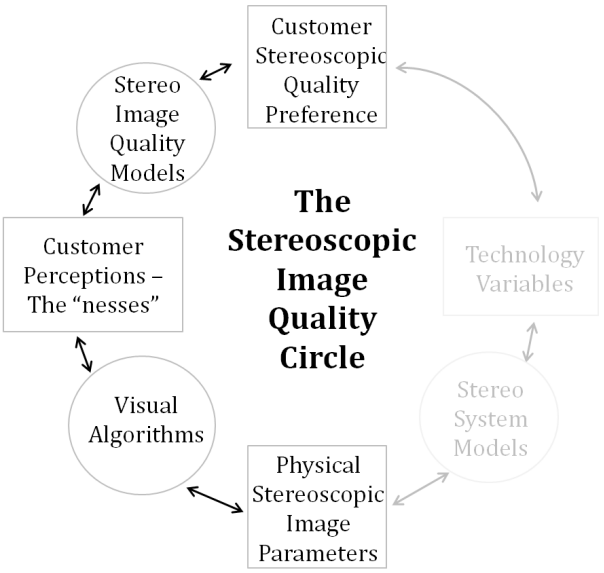


Figure 5.3: The Stereoscopic Image Quality Circle components we focused on in the present work.

Accordingly, less focus has been shown so far in the literature on the perceptual attributes that stay at the basis of the quality appreciation of stereoscopic images, as explained by Engeldrum. And, as seen in Section 4.8, only a limited number of studies have been performed that focused on the integration of different such perceptual attributes into a stereoscopic image quality model. Moreover, to our knowledge, no approach was proposed that tried to formalize the processes that lie behind the stereoscopic image quality assessment into a unitary framework.

5.2.1.2 The formal definition of the SIQC

The adaptation that we propose for the original IQC definitions [Engeldrum 2004b] to the stereoscopic context, *i.e.* to the SIQC, is the following:

Stereoscopic image = *a colorant arranged in a two-component (binocular) structure in order to convey* three-dimensional information *to a human observer, when watched on/through a special stereoscopic device.*

Stereoscopic image quality = *the integrated perception of the overall degree of excellence of a stereoscopic image* (in the *beauty-contest* sense).

Customer stereoscopic image quality rating (the stereoscopic image quality value) = *the judgment a customer renders for a sample stereoscopic image, independent of context, application, and fitness for purpose.*

Technology variables (the things we control) = *the items that are used to describe a stereoscopic imaging product, such as dots per inch or pixels per inch, image sensor megapixels, screen parallax, or other various visualization parameters.*

Customer perceptions (the *nesses*) = *the perceptual attributes, mostly visual, that form the basis of the stereoscopic quality preference or judgment by the customer.*

Physical stereoscopic image parameters = *quantitative and usually obtained by physically measuring the stereoscopic image with instruments or computations on a stereoscopic image file.*

Stereoscopic system/image models = *formulas, physical models, algorithms, or computer code that connect the physical stereoscopic image parameters and the technology variables.*

Visual algorithms = *formulas, models, or computer code that connect physical stereoscopic image parameters to customer perceptions; recipes that are used to compute a value of a* ness *from a physical stereoscopic image measurement.*

Stereoscopic image quality models = *link customer perceptions, the* nesses*, with customer stereoscopic image quality ratings. The stereoscopic image quality model inputs are values of* nesses *and the output is the customer stereoscopic image quality value. This is the ultimate destination in the many-to-one mapping process of the stereoscopic image quality circle: a one-number summary description of stereoscopic image quality.*

The rest of the observations we included in Section 4.9 related to the particularities of the IQC are compatible as-they-are to the new proposed SIQC.

5.2.2 The stereoscopic system

For being able to perform various tests relative to our research subject, we assembled a complete stereoscopic system, that would include a *stereoscopic camera* for capture, a standard *image processing* procedure, and two *stereoscopic displays* (a 3D TV screen and 3D glasses).

The detailed structure of our stereoscopic system will be given in Chapter 6, along with the precise geometrical model defining it.

5.2.3 The datasets

For the needs of our investigation on the human perception of the stereoscopic image quality, we have progressively created *three datasets*, by exclusively using our stereoscopic system.

This was an elaborated and time-consuming task, due to the precise criteria that we followed each time. However, the result was worth the effort, since the stereoscopic datasets obtained were each time compatible with the purpose of each subjective test for which we employed them.

The detailed description of each dataset that we created and also a preview of the images they include are given in Appendix A.

These datasets can also be found online, on the web site that we built, dedicated to the MOOV3D project.

5.2.4 The experimental studies

Given the implementation of the three previous steps – the elaboration of a framework, the set up of a stereoscopic system, and the creation of three stereoscopic datasets –, *experimental studies on stereoscopic images* represented the next natural step in our research project.

Such studies were a key element in our undertaking of better understanding the human judgment on stereoscopic images and the way this could be translated into a stereoscopic image quality model, important component of the SIQC.

The succession of the experiments we have implemented is illustrated in Figure 5.4 and justified by the following lines.

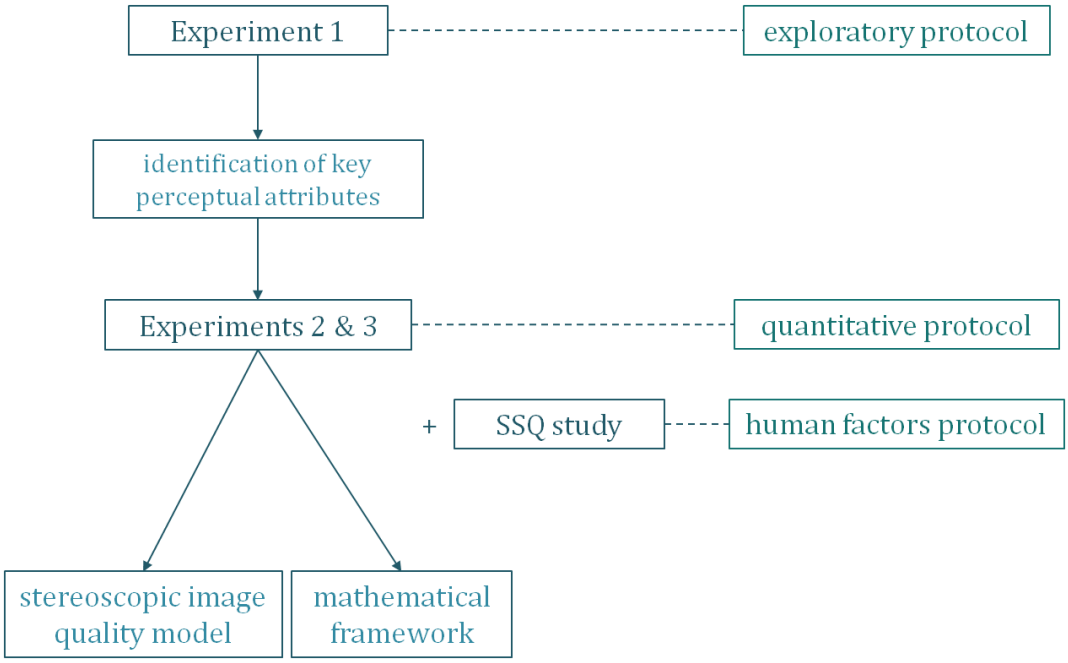


Figure 5.4: The succession of experiments performed.

Since the mechanisms underlying the stereoscopic image quality assessment are definitely not fully understood yet, we considered most appropriate to start our experimental chain with an exploratory study. As argued in Section 4.7.4, this type of investigation was the most suitable in our particular situation, where we intended to identify the main perceptual attributes related to quality on which the observers focus when watching stereoscopic images.

The results of this exploratory study set the stage for subsequent quantitative experiments. These succeeding studies only focused on the several perceptual attributes identified during the first experiment. They had the purpose of drawing conclusions on the importance of each of the perceptual attributes relative to the overall customer rating on stereoscopic image quality and to explain their relation with the physical parameters of the stereoscopic images. Consequently, they allowed to refine the first sketch of a stereoscopic image quality model obtained after the exploratory study.

We consider important to stress the fact that our focus was less on the precise vocabulary used to refer to each concept implied by the SIQC, but on the concepts themselves. We managed to do this by always explicitly defining the terms that we used during our experiments. Thus, we could avoid the bias due to the different interpretations that the participants would give to the notions used.

Also, it is essential to mention that our approach was a *quality approach*, opposed to the impairments approach, delimitation explained in Section 4.7.2. By this, we mean that we always asked our participants to evaluate the stereoscopic images displayed in an absolute sense, and not relative to references. We did this because our focus was on understanding how stereoscopic image quality could be defined by a naïve observer, when no a priori information or reference examples exist.

Conclusions

This text introduced the objectives that we set in our work. It explained our vision and the steps that we followed in our explorations on the stereoscopic image quality perceived.

The following chapters contribute to the overall view on our approach by first describing the stereoscopic system built in our laboratory. Then, the experiment protocols that we applied during our studies are illustrated in detail.

Chapter 6

Our Stereoscopic System

Introduction

For the various tests performed in the context of this work, a complete stereoscopic system was necessary. The most important criterion that influenced our choices in assembling such a system was the fact that the current research project is part of a wider approach that focuses on devices and applications for the consumer market. Therefore, the stereoscopic system built in the GIPSA laboratory was entirely composed of consumer devices or prototypes of such devices.

6.1 Capture

All the stereoscopic images that we used have been taken using the *compact Fujifilm FinePix Real 3D W3 camera* shown in Figure 6.1.

(a) Front view. (b) Back view.

Figure 6.1: The Fujifilm FinePix Real 3D W3 camera [Fujifilm FinePix REAL 3D W3 2013].

This dual-camera has a lens separation of 7.5 cm, slightly larger than the average interocular distance, and two CCD[1] sensors of the 1/2.3-inch-type, with a diagonal of 7.66 mm. A 3.5-inch

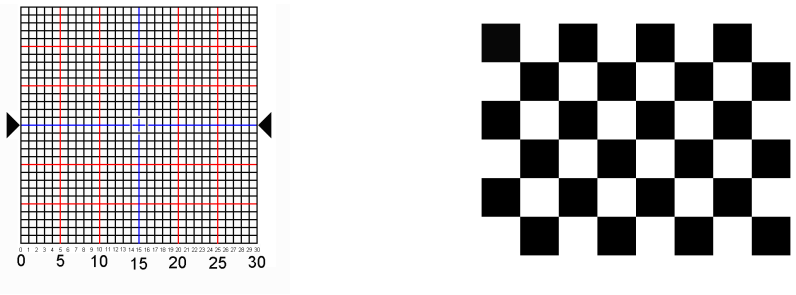
[1] charge-coupled device

auto-stereoscopic display of lenticular type is integrated on the back of the camera, as it can be seen in Figure 6.1b, and allows for live view and for the visualization of the recorded images.

Despite the availability of larger resolutions, all the images for our tests have been taken at *1080 p*. This mode, also called *full HD*, implies a resolution of 1920 × 1080 pixels (2.1 megapixels) and thus a widescreen aspect ratio of 16:9. We preferred this resolution mode for reasons of maximizing the compatibility with our display devices by eliminating the necessity of down-sampling.

The optical zoom of the camera allowed for focal length variations from 6.3 to 18.9 mm, equivalent to 35 to 105 mm on a 35 mm camera. This means that both wide and telephoto images could be recorded.

Since the exact geometry of the capture mechanisms in the camera was not publicly available, nor shared under request, we used black-box methods like measurements or calibration, in order to understand these underlying mechanisms and to compute the capture parameters. The measurement grid and calibration pattern used in our tests were those in Figure 6.2.



(a) Measurements grid (b) Calibration pattern

Figure 6.2: Schemas used for determining the camera parameters.

With a setup as the one illustrated in Figure 6.3, we were able to determine the field of view of the camera lenses [Jongerius 2012]. The measurements have been effectuated on the left lens, since the two lenses of the camera have been considered to have very close to identical optical properties. The distances x, y, and h have therefore been precisely measured after making red marks on the grids at the extreme points that would fit in the recorded images, and simple geometrical calculations allowed determining the field of view of the lens, which is the double of the α angle. In consequence, the results showed that the field of view of the Fujifilm W3 camera lens varies from about 17.3 ° for the 18.9 mm focal length to approximately 46.4 ° for the 6.3 mm focal length.

After several other tests on images using grids like the one in Figure 6.2a we made the hypothesis of *a parallel configuration* of the two lenses of the compact camera.

Furthermore, the calibration of the camera was performed separately for two fixed focal lengths (6.3 mm and 18.9 mm), by analyzing each time a set of images of the calibration pattern in Figure 6.2b with the Camera Calibration Toolbox for Matlab [Bouguet 2010]. Each time, the calibration indicated a horizontal displacement of both the left and the right views on the corresponding sensors relative to their centers. This displacement is called *sensor axial offset* and, for the parallel camera configurations, indicates the distance by which the center of each

imaging sensor has been moved away (outwards) from the optical axis of the lens to achieve convergence [Woods 1993]. The precise sensor axial offset in the case of our stereoscopic system is given by the p_{xl}, p_{xr}, p_{yl}, and p_{yr} parameters, and its illustration on the sensor in the x direction can be seen in Figure 6.6.

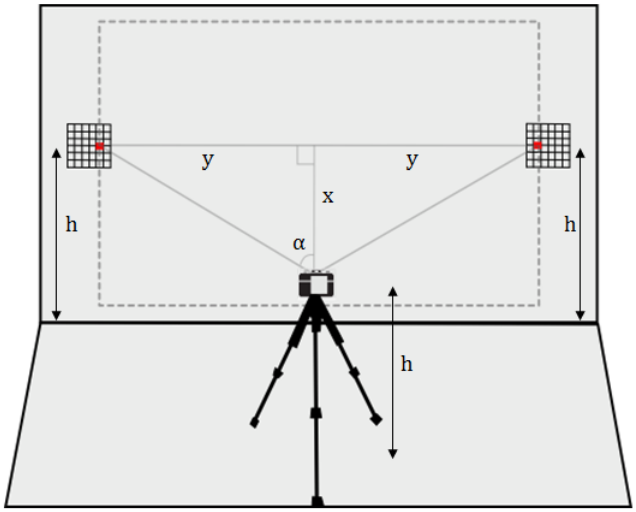


Figure 6.3: Set up for measuring the camera field of view [Jongerius 2012].

Therefore, the geometrical structure of the capture mechanisms was drawn as presented later in Figure 6.6 and this model was later confirmed by our verification tests.

6.2 Image processing

The 3D images rendered by the camera often show a supplementary horizontal reciprocal shift of the left and right views. This displacement is generated by the automatic parallax configuration of the camera, such that the recorded disparities would be each time adjusted in function of the object in focus at the moment of the capture, object that would thus be rendered at screen level. The displacement also implies a variable reduction of horizontal resolution, due to the reduced surface on which the left and the right views superpose, after being relatively shifted horizontally.

However, the stereoscopic images taken by the Fujifilm W3 3D camera are saved as *.mpo files* and this format does not store the two views of a stereoscopic image in their trimmed form, but stores the original full-resolution left and right views and, with these, the horizontal shift as a parameter.

Thus, during the processing phase, we could always set the shift parameter to zero and extract the original full-size left and right views from the stereoscopic files. This allowed us to maintain a unique geometrical model for all our test images. In our model, the zero-disparity plane to be represented at screen level would always be situated at a constant distance in front of the camera for a given focal length.

6.3 Displays

Two display technologies were used for the visualization of the stereoscopic content that we produced.

The first display was a *Panasonic TC-P50VT20 stereoscopic TV screen* with plasma technology, shown in Figure 6.4. The size of this screen is of 111 cm × 70 cm and its resolution is 1080 p. In 3D mode, this display works in synchronization with the active shutter glasses shown also in Figure 6.4.

Figure 6.4: The stereoscopic Panasonic TV screen and its shutter glasses [Sony TX-P50VT20B 3D Plasma TV 2013].

The second display was a *prototype head-mounted device*, in the form of a pair of *3D glasses with two micro OLED displays*. This device, that can be seen in Figure 6.5, did not include a head-tracking option, therefore we consider the concept of *personal display system* (or *personal viewing system*) [Howarth 1997] compatible to define it.

Figure 6.5: The prototype 3D glasses with micro OLED displays [MicroOLED 2013].

The field of view of the micro screens of the prototype 3D glasses is of approximately 28 ° and their resolution is WVGA[2], equivalent to 854 × 480 pixels. Inside the glasses, a system of lenses is used for the magnification of the image on the two screens. Also, optical adjustments by each participant are necessary, in order to adapt the glasses to the different diopters values of the observers.

The maximum resolution accepted for the content displayed on the prototype 3D glasses was 720 p in side-by-side mode, therefore our images were adapted to this format when displayed on

[2]wide video graphics array

this device. This device was a prototype produced within the MOOV3D project.

In order to display our stereoscopic images on any of the two displays, a 3D compatible computer was also used, which was connected to the displays by HDMI[3] 1.4 cables.

6.4 The geometry of our stereoscopic system

Given these devices, our stereoscopic system covers all the necessary phases from capture to visualization. Its precise modeling is an important part of our work and is essential in order for us to be able to perform tests with the controlled variation of the desired parameters.

The system composed of the Fujifillm W3 camera and the Panasonic TV display could thus be geometrically defined by a series of mathematical relations between the four phases of our stereoscopic chain:

(1) the choice of the real-world scene to capture;

(2) the recording of the captured image on the dual-sensor;

(3) the rendering of the recorded image on the stereoscopic screen;

(4) the visualization of the displayed image by an observer.

For this, we considered that a real-world point $P_o(x_o, y_o, z_o)$ photographed with our camera during phase (1) would correspond to a point on each camera sensor, *i.e.* to point $P_{cl}(x_{cl}, y_{cl}, z_{cl})$ on the left sensor and to point $P_{cr}(x_{cr}, y_{cr}, z_{cr})$ on the right sensor during phase (2). The sensor points would be then rendered as the screen-points $P_{sl}(x_{sl}, y_{sl}, z_{sl})$ and $P_{sr}(x_{sr}, y_{sr}, z_{sr})$ on the Panasonic TV during phase (3), then as the image point $P_i(x_i, y_i, z_i)$ perceived by an observer in the visualization space during phase (4). The four sets of Cartesian coordinate systems considered during each phase have the following origins: (1) O_o at the middle of the camera-base distance, (2) O_{cl} and O_{cr} in the centers of the two sensors, (3) O_s in the center of the screen, and (4) O_i at the middle of the interocular distance.

First, the transition between phases (1) and (2) is geometrically illustrated in Figure 6.6 (inspired from [Woods 1993]), where t is the notation for the baseline of our stereoscopic camera, f represents the considered focal distance, w_c is the sensor width determined by calibration ($w_c = 1920\,px \approx 5.57\,mm$), and p_{xl} and p_{xr} represent the horizontal axial offsets of the left and right sensors, determined by calibration as well, of the values given in Table 6.1.

The formulas that detail this transition are the following:

$$\begin{aligned} x_{cl} &= \frac{f \cdot (2x_o + t)}{2z_o} + p_{xl}, \qquad y_{cl} = \frac{f \cdot y_o}{z_o} + p_{yl}, \\ x_{cr} &= \frac{f \cdot (2x_o - t)}{2z_o} + p_{xr}, \qquad y_{cr} = \frac{f \cdot y_o}{z_o} + p_{yr}, \end{aligned} \tag{6.1}$$

[3]high-definition multimedia interface

Second, the correspondence between the sensor points and the screen points, *i.e.* between steps (2) and (3), is defined as a scaling with factor M equal to:

$$M = \frac{w_s}{w_c}, \tag{6.2}$$

where w_s and w_c are the screen width and the sensor width.

Table 6.1: The values of the p_{xl}, p_{xr}, p_{yl}, and p_{yr} parameters computed by calibration for the Fujifilm FinePix Real 3D W3 camera configuration.

	in mm				in px			
focal distance	p_{xl}	p_{xr}	p_{yl}	p_{yr}	p_{xl}	p_{xr}	p_{yl}	p_{yr}
$6.3\,mm$	-0.191	0.034	0.145	0.11	-66.187	11.884	50.707	38.55
$18.9\,mm$	0.278	0.312	0.027	-0.008	95.185	106.899	9.297	-2.632

This amounts to the following mathematical relations:

$$\begin{aligned} x_{sl} &= M \cdot x_{cl}, \qquad y_{sl} = M \cdot y_{cl}, \\ x_{sr} &= M \cdot x_{cr}, \qquad y_{sr} = M \cdot y_{cr}, \end{aligned} \tag{6.3}$$

Last in the stereoscopic chain, the geometrical representation of the transition from phase (3) to phase (4) is given in Figure 6.7, where e is the interocular distance of the observer and D is the visualization distance. The case considered is that of a point represented behind the screen. For the cases where the image points are perceived in front of the screen, a similar judgment applies.

This schema can be translated in the following mathematical formulas:

$$\begin{aligned} x_i &= \frac{e \cdot (x_{sl} + x_{sr})}{2 \cdot (e + x_{sl} - x_{sr})}, \\ y_i &= \frac{e \cdot (y_{sl} + y_{sr})}{2 \cdot (e + x_{sl} - x_{sr})}, \\ z_i &= \frac{e \cdot D}{e + x_{sl} - x_{sr}}. \end{aligned} \tag{6.4}$$

The last equation, that details the depth z_i perceived by the observer, can be written as compatible with the theoretical Equation 3.1 on perceived depth in the following way:

$$z_i = \frac{(x_{sr} - x_{sl}) \cdot D}{e - (x_{sr} - x_{sl})} + D, \tag{6.5}$$

where $x_{sr} - x_{sl}$ represents the screen disparity d. This form is justified by the fact that

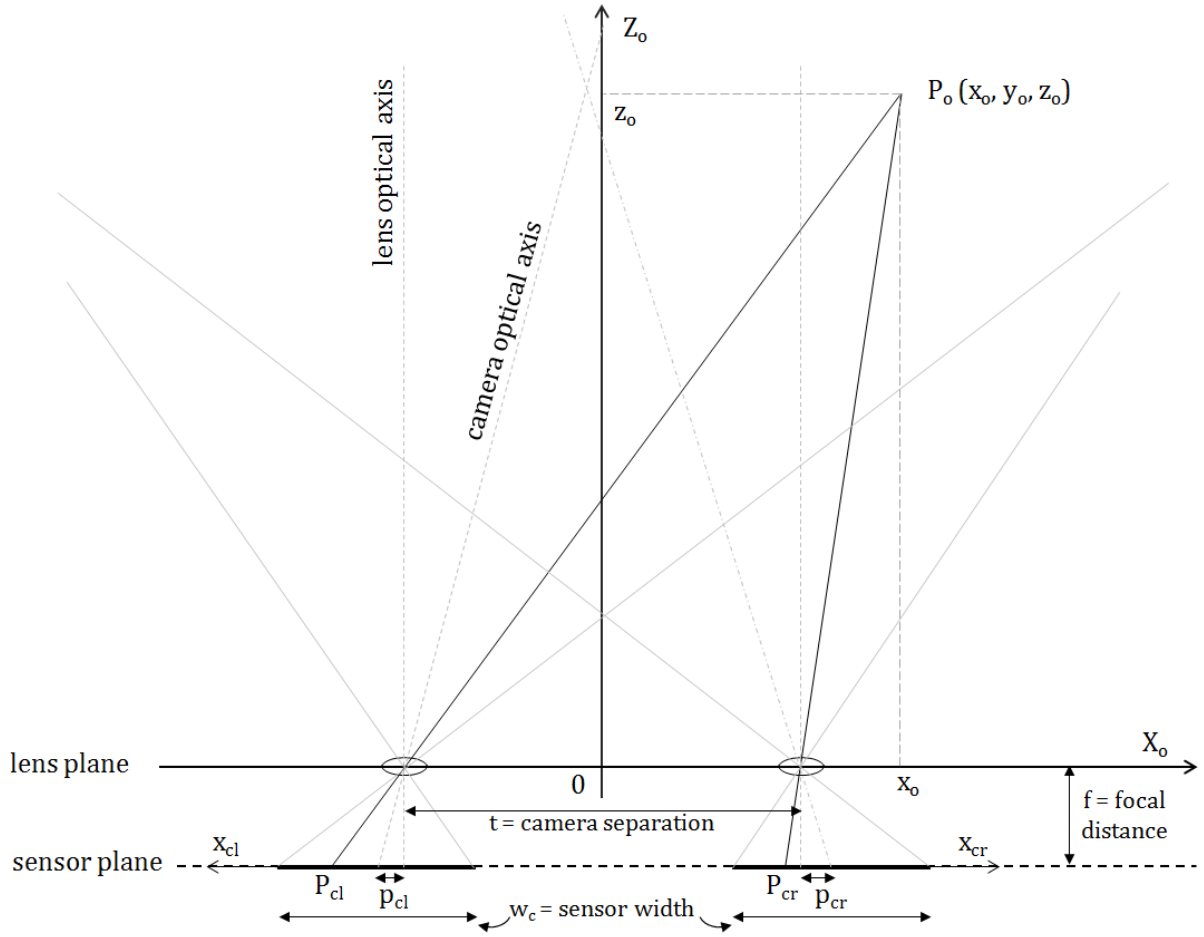


Figure 6.6: The geometrical representation of the capture mechanisms of the considered dual-camera.

the depth at which the observer perceives a represented point (z_i) equals the depth perceived relative to the screen (Equation 3.1) added to the distance between the observer and the screen (D).

All these expressions lead to the general model of correspondence between a real-world point $P_o(x_o, y_o, z_o)$ and the image point $P_i(x_i, y_i, z_i)$ perceived by an observer when using our system both for capture and visualization:

$$\begin{aligned} x_i &= \frac{e \cdot M \cdot (2x_o \cdot f + z_o \cdot (p_{xl} + p_{xr}))}{2 \cdot (e \cdot z_o + M \cdot f \cdot t + M \cdot z_o \cdot (p_{xl} - p_{xr}))}, \\ y_i &= \frac{e \cdot M \cdot (2y_o \cdot f + z_o \cdot (p_{yl} + p_{yr}))}{2 \cdot (e \cdot z_o + M \cdot f \cdot t + M \cdot z_o \cdot (p_{xl} - p_{xr}))}, \\ z_i &= \frac{e \cdot D \cdot z_o}{e \cdot z_o + M \cdot f \cdot t + M \cdot z_o \cdot (p_{xl} - p_{xr})}. \end{aligned} \tag{6.6}$$

Once this model defined, a suggestive representation can be drawn: the simulation of the deformations that real-world objects might undergo when using our stereoscopic system from

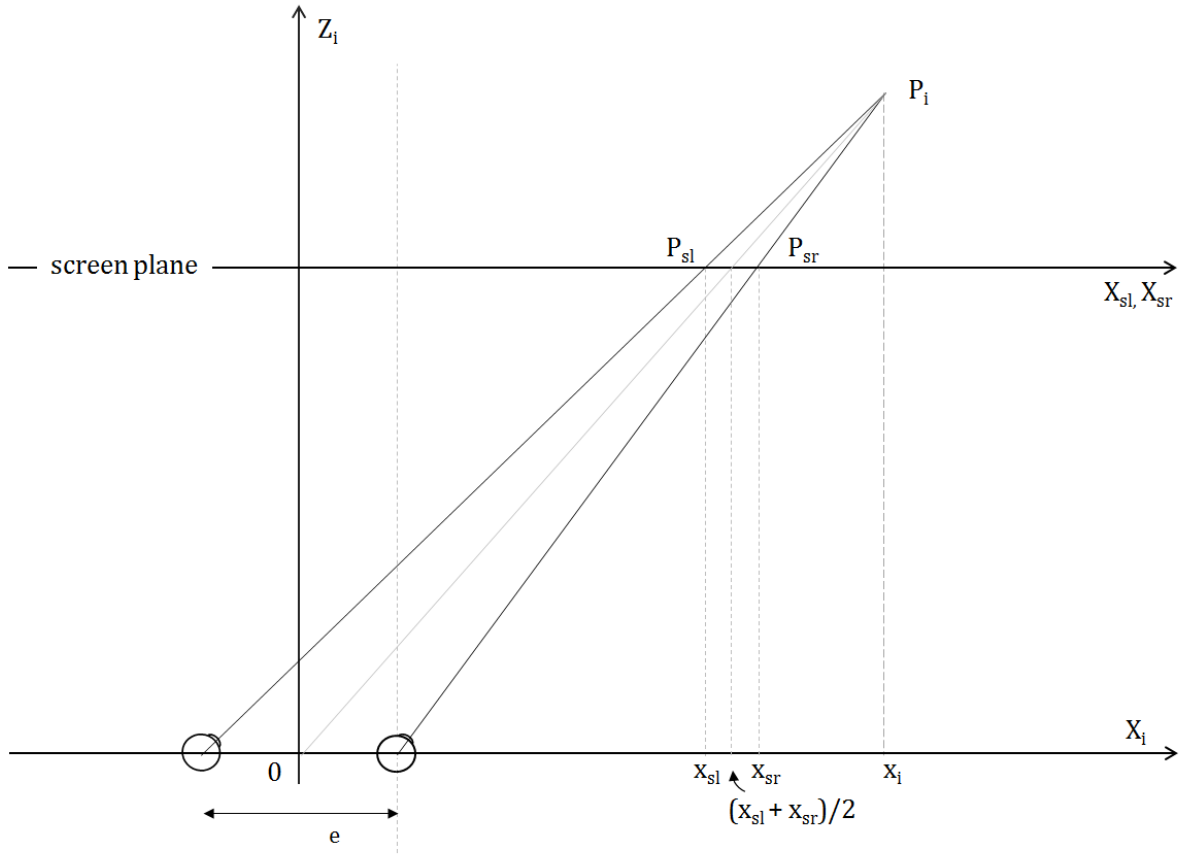


Figure 6.7: The geometrical representation of the depths perception on the considered stereoscopic screen.

the capture to the visualization phase. Figure 6.8 shows these distortions on a set of equidistant real-world points placed at $100\,mm$ one from another, that would be captured and rendered with our stereoscopic system, given a visualization distance $D = 1.5\,m$ and an interocular distance $e = 65\,mm$. Figure 6.9 shows the same situation, but for a changed visualization distance $D = 3\,m$.

Other variations in the parameters of this system would generate other distortions in the perceived content. However, by taking into account the observations in [Woods 1993], we can notice that the depiction of the reproduced space for our system, in the form in which it is presented here, is satisfactory, with the exception of a compression of the perceived depth augmenting with the distance to the photographed object.

Unfortunately, a similar detailed definition of the system composed of the Fujifilm camera and the prototype 3D glasses was not possible. For each participant, the configuration of the 3D glasses is different due to the manipulation of the lens focus that he or she makes in order to adapt them to their view. Moreover, the distance from the glasses to the eyes of the observers is always different because of variations in the facial features of the viewers, that make the 3D glasses screens to be placed closer or further from their eyes, without us being able to measure precisely these distances.

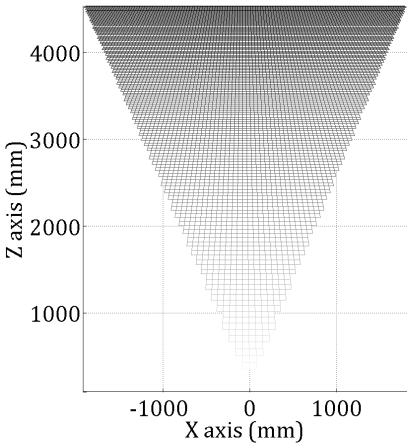


Figure 6.8: The geometrical distortions produced by our stereoscopic system for a visualization distance of $1.5\,m$.

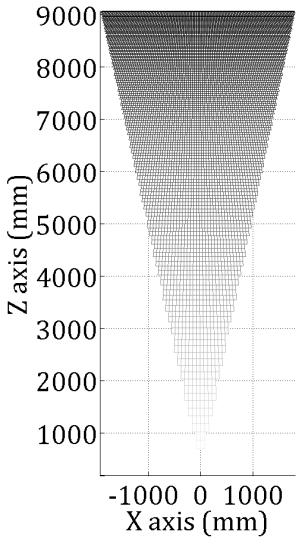


Figure 6.9: The geometrical distortions produced by our stereoscopic system for a visualization distance of $3\,m$.

Thus, in the experimental studies that we present in the following chapters, we will limit the very technical observations only to the case of stereoscopic viewing using the TV screen and express our observations related to the 3D glasses only in more general, less mathematical, terms.

Conclusions

Given these descriptions, observations, formulas, and graphical representations, our stereoscopic system is complete and also entirely schematized. This makes it available to be exploited in the various experimental studies that follow.

Chapter 7

Our Experimental Protocols

Introduction

The present chapter contains the detailed description of the experimental procedures that we used in our studies, in which we explored how a stereoscopic image quality model could be designed.

Three *subjective assessment protocols* have been used to this end. The first one was adapted for an *exploratory study* aimed at the identification of the perceptual attributes that lay at the basis of the human stereoscopic image quality judgment. The second protocol was used for several *quantitative studies* designed to measure the influence of each of the identified perceptual attributes on the overall judgment on stereoscopic image quality. And the third protocol was used during a part of our quantitative tests for the collection of data related to the *visual discomfort* symptoms.

7.1 General considerations

Despite the different specificity of each of the experimental approaches, a series of aspects are common to all the protocols that we used. These are presented in the following.

7.1.1 Participants

The selection of the participants to our experiments has been attentively performed.

First, all the participants were *over 18 years old*, fluent in *French or English*, and chosen to be *naïve* from the point of view of the familiarity with the subject under study, *i.e.* none of them was directly involved in activities related to the perceived quality of stereoscopic images. The participants have been notified about the experiment mostly through mailing lists or paper notices in, or in contexts related to, academic institutions. Therefore, the vast majority of the participants were university students, doctoral or post-doctoral students, researchers, or other academics. However, our population samples were not limited to these categories.

Second, all the participants gave their signed consent related to their voluntary involvement in the experimental studies. Also, for each participant, a *personal non-nominal data file* was recorded, containing statistical data related to their age, gender, activity, familiarity with the stereoscopic technologies, *etc.* These personal data have been processed in complete accordance with the French *Act on Data-processing, Data Files, and Individual Liberties* of January 6, 1978 (*Loi Informatique et Libertés*).

Third, all the persons interested to participate in the experiment were screened with *vision tests* and only those that successfully passed all the tests were kept as valid participants. The first test evaluated the *visual acuity* of the participants and was performed using the *Snellen chart*. Only the persons that had a vision of 20/25 or better were kept, considering that the normal acuity reference is 20/20. The meaning of the 20/25 visual acuity is that the subject can read the chart from 6 m (20 feet) as well as a normal person could read the same chart from 12 m (40 feet) away. The second test evaluated the *color vision* and for implementing it we used a selection of the *Ishihara plates*. All the selected illustrations needed to be correctly read in order for the test to be validated. The third factor to be assessed was the *stereoscopic acuity*, *i.e.* the acuity of the stereoscopic vision. For this, two solutions were employed, either an *online test* made available by the McGill University[1], or the *TNO test*. According to the TNO test recommendations, we considered as valid the participants who had a stereoscopic acuity of at least 60 seconds of arc, while the TNO test allowed measuring stereoscopic acuity values from 15 seconds of arc to 480 seconds of arc.

Last, for all the experiments carried out in the context of the present work, except for the exploratory study, a *reward* was offered to the participants at the end of their intervention. This was each time in the form of a USB drive.

7.1.2 Experimental environment

All the experiment sessions during which the participants visualized stereoscopic images were implemented in a *home cinema* simulated environment. In this manner, the two experiment rooms we used had dark grey walls and no windows. A daylight lamp of 25 W was used to light these two rooms of approximately 12 m^2 and 16 m^2 in surface, respectively. Only one participant performed a given task at a time.

[1] http://3d.mcgill.ca/

7.1.3 Stimuli

The stimuli sets visualized by the participants to our experiments were the three datasets that we created. These were the *MOOV3D - GIPSA Datasets 1, 2, and 3*. Their full description can be found in Appendix A. Also, the exact dataset that was used in each experiment and the justifications of their choice will be given later, in the chapters describing each one of the experiments, *i.e.* Chapters 8, 9, and 10.

7.1.4 Structure

The general structure of each of our experimental studies was made up of four main parts:

(1) the *vision screening* of the participants,

(2) their *familiarization with the task*,

(3) the *assessment experiment* itself, and

(4) the *collection of statistical data* on the participants.

The *vision screening* was performed with the help of standardized vision tests, as explained in one of the previous paragraphs, in Section 7.1.1.

Concerning the experiment task, in order for the participants to understand it and to become familiar to it, a *task sheet* was printed for each different experiment. All the task sheets of our experiments are reproduced in Appendix D. These included an overview of the steps of each experiment, the actions that the participant had to perform, and also the definitions of the notions that might be ambiguous to naïve participants. Before starting an evaluation experiment, the participant had to read the corresponding task sheet. Since the same document was presented to all the participants to the same experiment, the variations in the presentation of the task were avoided, as it might happen when giving oral instructions.

The development of the *assessment trials* will be discussed for each case separately, in the following sections, since it varies considerably in function of the protocol used.

The *collection of statistical data* was performed either with online forms or on paper, but always at the end of the experiments, in order to avoid fatiguing the participant with secondary tasks before the visualization experiment.

7.1.5 Test plan

For each study, we prepared a detailed test plan with all the steps to be followed by the experimenter, just as explained in Section 4.7.3.6. The test plan of each experiment will be referred to later in this manuscript. These scripts are included in Appendix D.

7.1.6 Implementation and data collection

All our experiments have been managed by the Soft Eye software application [Ionescu 2009] available in our laboratory. The application allowed us to configure the random succession of the test stimuli, to set their display duration, to configure the messages or the content to be displayed, to store the participant ratings when needed, *etc.*

The data collection has been performed differently in function of the specificity of each protocol and it will be discussed for each case in particular.

7.1.7 Data processing

The data processing was performed each time with the Matlab numerical computing software. A series of different statistical procedures were employed in order to process the brute results obtained for each experimental study. All these methods are explained in Appendix B.

7.2 The exploratory protocol

Since the work described in this manuscript is related to a field in which standardized models of quality assessment do not exist yet, we consider that the exploratory studies, with their role of pathfinders, as described in Section 4.7.4, have at present a critical importance.

This is the reason why we started our investigation on the modeling of the stereoscopic image quality by performing such an exploratory study. We had the purpose of highlighting thus the aspects on which the human observers focus most when watching stereoscopic images and judging their quality, then concentrated on these aspects more in detail, with the help of more formal quantitative tests. The exploratory study we implemented consisted of two parts.

The first part was a *classification* step, where the participants validated our subjective view on the structure of the MOOV3D – GIPSA Stereoscopic Dataset 1 and confirmed the desired variability of its content.

The second part was a *visualization and verbalization* step, where the viewers gave their opinions orally on the quality of a selection of stereoscopic images they watched.

The two parts were complemented, as described in Section 7.1.4, by a testing stage before the visualization step, which allowed eliminating the participants with visual deficiencies, and also by a questionnaire at the end, which allowed gathering demographic data.

7.2.1 Part I - Database validation

Because all the images in Dataset 1 had been photographed by only two persons, the structuring of the collection of stereoscopic images according to the criteria explained in Appendix A.1 could be biased. This is why during the first part of our study we used subjective testing with multiple participants in order to confirm or invalidate the good distribution of the images in the

18 categories and to facilitate the subsequent selection of test material for the second and the most important part of our test.

To this purpose, the left views of all the 158 stereoscopic images in the database were printed on cardboard paper (approximately A5 in size). Each of the participants who took part in this test was offered the entire collection of printed images and a task sheet in which they were asked to classify the images (task sheet reproduced in Appendix D.1). This sheet was either in English or in French, in function of the language the participant was more comfortable with. The schema in Figure A.2, representing the considered dataset structure, was displayed at large scale on a table in the experiments room. In this fashion, the participants were supposed to do the classification by placing each printed image on the table, in the case corresponding to the category chosen for it.

An example of classification of the given collection of cardboard images is illustrated by Figure 7.1.

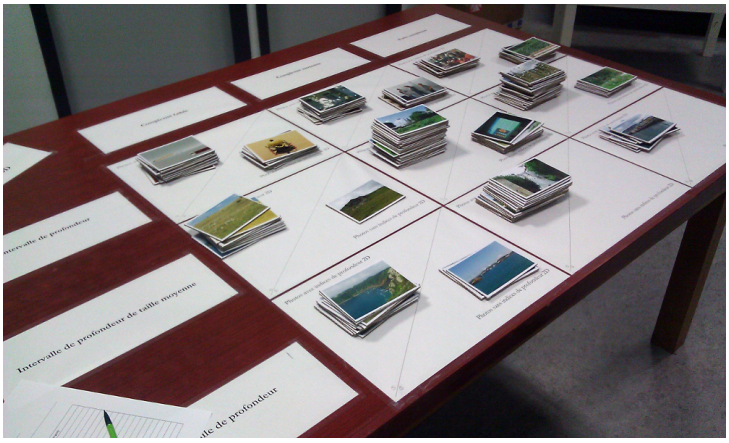


Figure 7.1: The experiments room during part I of the exploratory study.

For the classification task, the participants had no time constraints. The task was considered accomplished when all the printed images were distributed in the cases of the given table.

7.2.2 Part II - Qualitative exploration of selected 3D content

For the visualization experiment, we used the Panasonic stereoscopic display with active shutter glasses described in Section 6.3. The television screen was situated on a table in front of the participant, who was advised to change the seat height for an optimal position, centered relative to the screen. The visualization distance was of 140 cm and it was chosen as a compromise between the ITU Recommendation for assessing stereoscopic pictures in force at the time [ITU 2008] and the television specifications. The experiment environment was configured as described in Section 7.1.2.

The qualitative exploration was carried out in the form of a session of visualization of 24 stereoscopic images in random order, which had been previously selected from the larger MOOV3D – GIPSA Stereoscopic Dataset 1, according to the results of the classification performed after Part I of this protocol. During the visualization, the participants were supposed

to freely express their opinions on the perceived quality and on the comfort/discomfort induced by the stereoscopic content displayed.

The experiment was thus a qualitative exploration and followed the principles of a semi-structured interview, procedure described in Section 4.7.4. The same explanatory text was offered to all the participants, either in English or in French, mentioning two main dimensions on which they were supposed to express their opinions, respectively *quality* and *comfort*. After each participant finished verbalizing his or her first impressions, extra questions were asked in order to clarify these first impressions and to make sure that the factors that were possibly considered irrelevant by the test subjects were also discussed. The supporting questions also allowed to maintain the focus of the observer on elements directly related to the perceived stereoscopic image quality and to avoid comments unrelated to the purpose of the experiment.

No time limit was set for the verbalization experiment neither. The participants were free to speak for as long as they wanted about how they perceived a given stereoscopic image. After the utilization of this protocol for our qualitative study, we could calculate a mean of a little over 1 minute per visualized image over all the participants.

The visualization test was performed either in English or French, in function of the preferred language of the participants, and the remarks that were made were recorded in order to be listened and processed later.

7.3 The quantitative protocol

Once our exploratory study results were analyzed (topic discussed in the next chapter, Section 8.4.2), a *validation* and, in the mean time, a *refinement* of these first results were necessary. As illustrated in our presentation on existing subjective assessment methods, in Section 4.7, the quantitative studies are the most suitable for this purpose. However, numerous quantitative protocols exist and we had to make a choice that would be adapted to our context.

First of all, in our work, the focus was on the stereoscopic image quality in its absolute sense, when no reference stereoscopic images exist, and when the observer is not constrained to make a judgment relative to a particular context of utilization, but a judgment of quality that is as general as possible. Therefore, we naturally made our choice among the *single stimulus* methods, no direct comparison between two different stimuli being of interest for us.

Next, we found ourselves in the situation where the effects of the interactions between several independent factors on the studied attributes were of interest as well, along with their individual effects, as explained in Sections A.2 and A.3, when detailing the construction of our stimuli test conditions. This led to large datasets to be analyzed quantitatively, hence we were constrained to choose a method that would be *fast and simple*, in order to compensate the heavy load given by the very large number of images to be assessed.

Given these considerations, the most appropriate experimental method to be used for our quantitative tests proved to be the *single stimulus method (SS)*, also referred to as the *absolute category rating (ACR)* method. Our choice was reinforced by results showing that this method, besides seven other under test, ensured the shortest total assessment time, proved the best level of ease of evaluation from the participant perspective, and presented a strong statistical

reliability [Tominaga 2010].

Consequently, our quantitative protocol followed the principles presented in Section 4.7.3.4, in the paragraphs on the SS method.

Concerning the particularities of our implementation, we fixed the stimulus display duration to 6 seconds, the pre-exposure mid-grey field display duration to 2 seconds, and we left the voting time of unlimited duration, in the sense that a new stimulus to be evaluated would be displayed as soon as the previous stimulus was rated. Calculations related to the experiment total time showed that, despite the unlimited time they had to their disposal, the majority of our participants expressed their votes in the first second after visualizing the stimuli. Also, as a post-exposure field, we showed the text "Vote now!", along with a recall of the possibles ratings from which the participant had to choose, those illustrated in Figure 4.11a. The exact structure of our trial assessments is shown in Figure 7.2.

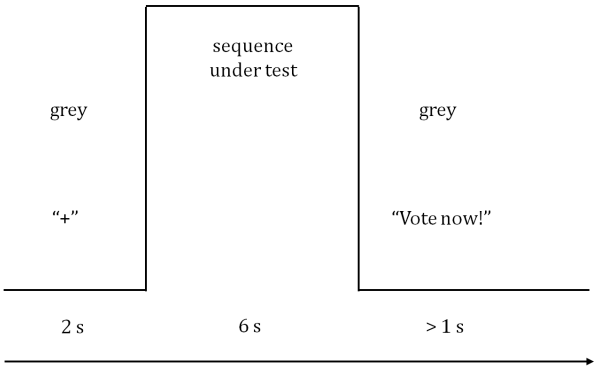


Figure 7.2: The structure of a trial assessment for the SS method, as we adapted it for our quantitative protocol.

Another particularity of implementation is related to the recording of the votes of the participants. The ratings from 1 to 5 were highlighted on a classical computer keyboard. In this manner, when the rating scale was displayed on screen during the voting phase, the participants could push the desired button, making the procedure simple and rapid. The rating selected was stored by the experiment software mentioned in Section 7.1.6 and extracted subsequently during the data processing phase.

Concerning the rating scale, we preferred the use of a scale which was both adjectival and numeric: 1 - bad, 2 - poor, 3 - fair, 4 - good, 5 - excellent. With such a scale, on one side, the perceived intensities to judge are named and, on the other side, an equal spacing between consecutive words on the scale is ensured.

The last modification that we brought to the classical SS implementation is related to the number of presentations of the same stimulus during the experiment. The original SS method suggests either one or three repetitions. In the case of our experiments, we were constrained, on one hand, by the large number of stimuli that would imply a long experiment duration, and, on the other hand, by the difficulty of rating a relatively new type of content which would imply a slow adaptation to the stimuli and to the rating strategy. As a consequence, we made the compromise of showing the stimuli set twice in our quantitative protocol. The first repetition would allow the participants to become familiar with the stimuli and with the rating task and

would eventually be discarded from the computation of the final results. The second repetition would be the only one kept for the analysis, since the rating strategy of each participant would be stabilized by then. Both repetitions however would participate to the screening procedure presented in Appendix B.3.

This quantitative protocol was equally adapted both for use in our experiments performed on the 3DTV screen described in Section 6.3 and in our experiments on the 3D glasses presented in the same section.

The visualization conditions were the same as for the visualization part of the exploratory study, with the exception that, for the experiments performed on the TV screen, the visualization distance was adapted to 150 m. This distance was subsequently a key parameter in the Matlab program that reproduced the stereoscopic system geometry model described in Section 6.4, thus finalizing the definition of all the parameters of our model.

The quantitative tests that we carried out following this protocol used Datasets 2 and 3 as stimuli.

7.4 The human factors protocol

Additionally to the exploratory and quantitative studies, a human factors protocol was implemented as well. It had the purpose of evaluating the simulator sickness symptoms induced by the visualization of stereoscopic content and of observing if these symptoms are influenced by the display used, by the task given to the observer, or by the particularities of the content displayed.

Such a protocol has the advantage that it can be easily implemented as integrated with any type of other subjective assessment protocol. This allowed us to perform it in the same time as some of our quantitative studies.

Central to the protocol was the *Simulator Sickness Questionnaire (SSQ)* introduced in Section 4.7.6. The original form in which it was initially proposed is shown in Appendix C. In our protocol, it was used in an adapted form.

Thus, from the set of 16 questions originally included in the test, we only kept the 12 that we considered fit for the context of visualizing static stimuli. These questions are illustrated in Table 7.1, along with their weights relative to each symptom cluster assessed: N (nausea), O (oculomotor), and D (disorientation). This modification did not reduce the correctness of the interpretation of our results, since the reasoning on the data collected was later done mostly as a comparative analysis of results obtained in the same conditions.

The original purpose of the use of the SSQ was to compare a large number of simulators from the point of view of the symptoms they generate, thus the authors of the questionnaire recommended to administer it only after the visual immersion period. In our case, the purpose was to evaluate how the human factors evolve from before to after the visualization of images on a given stereoscopic display, therefore we complemented the post-exposure data collection with a pre-exposure application of the questionnaire, thus adapting its utilization to our experimental context.

The scale used for each question was the same as in the original questionnaire: a 4-point

adjectival scale using the keywords: none (score 0), slight (score 1), moderate (score 2), and severe (score 3).

Table 7.1: The adapted SSQ used in our study (the questions used and the weights for computing the N, O, and D symptom scores).

	Weight		
SSQ Symptom	N	O	D
General discomfort	1	1	0
Fatigue	0	1	0
Headache	0	1	0
Eyestrain	0	1	0
Difficulty focusing	0	1	1
Nausea	1	0	1
Difficulty concentrating	1	1	0
Fullness of head	0	0	1
Blurred vision	0	1	1
Dizzy (eyes open)	0	0	1
Dizzy (eyes closed)	0	0	1
Vertigo	0	0	1

Conclusions

All the experimental protocols presented so far constituted rigorous and precise frameworks for our investigations. They allowed the implementation of several experimental studies that were rich in results. The rest of this manuscript concentrates on the results of these studies, presenting them in detail.

Part III

Results

Chapter 8

A Qualitative Study

Introduction

The first step in our investigation on the human perception of the stereoscopic image quality was *an exploratory study*. It was implemented in order to highlight the main elements on which human observers center their judgment of quality when presented with stereoscopic images, and it allowed us to draw a first sketch of a stereoscopic image quality model.

In order to increase the precision of such an exploratory study, which is characterized by an important amount of subjectivity, the experiment included a preliminary test on the stimuli set to be used and was in consequence structured in two parts. The first part was a subjective *classification* step performed on the stimuli, while the second part of the exploratory study was a *visualization and verbalization* step that used the stimuli set selected after the first part.

The following sections give comprehensive details on the processing of all the data collected during this study and on the results obtained. These information are illustrated in the order in which the experiment was performed, *i.e.* first the results of the stimuli classification and the subsequent stimuli selection, then the results of the visualization and verbalization test.

8.1 Experimental protocol and participants

This study followed closely the *exploratory protocol* described in Section 7.2 for each of its two parts and used as initial stimuli set the 158 images in the *MOOV3D – GIPSA Stereoscopic Dataset 1* illustrated in Appendix A.1.

A total of 27 persons, aged between 21 and 43, with an average age of 29 and a median age of 27, and mostly male (70 %), participated in the experiment. Among these participants, three persons participated only in the classification part because of their incapacity of perceiving stereoscopic content or of their insufficiently precise stereoscopic acuity, and three persons participated only in the visualization part. This led to a total of 24 persons for each of the two parts of the test.

Only two of the participants owned a stereoscopic device, while the contact of the majority of the participants with stereoscopic content in general was only occasional. Two of the participants had never watched stereoscopic images or videos before.

The large majority of the participants were highly qualified, with 5 undergraduate students, 11 graduate students, 9 highly qualified research staff, and only 2 persons with no higher studies.

8.2 Part I – Classification

8.2.1 Data collected

After each participant performed the classification of the given stimuli in the 18 categories, the results of this classification were directly stored in the form of a correspondence between the ordinal number of each printed image and the category that was allocated to it. The totality of results of all the participants thus led to a matrix storing every category that had been allocated to each stimulus, *i.e.* a matrix of size 158×24.

8.2.2 Data processing and results

In order to analyze the classification data, the matrix containing the raw classification results was used to obtain 158 histograms, one for each image. In these histograms, the height of each bar corresponds to the number of participants who have classified the image in the category given by the ordinal number of the bar. An example of such histogram can be seen in Figure 8.1. We will refer to this representation also by *histogram of votes*, since it represents the number of *votes* that each category has accumulated for a certain image.

For finding the representative category to which a certain image was subjectively judged to belong, the most obvious approach was to prefer the category corresponding to the peak of its histogram of votes as the solution. The examples in Figure 8.2 show that this was a basic but intuitive approach, suitable in the majority of cases.

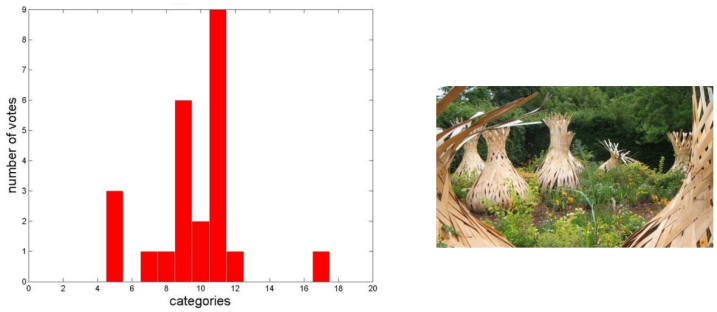


Figure 8.1: An example of a category histogram (or histogram of votes), corresponding to image n^o 15.

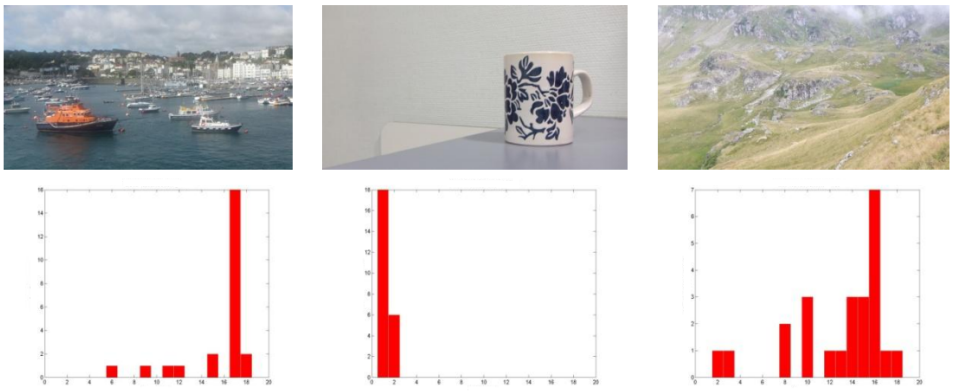


Figure 8.2: Examples of images that can be naturally classified in the category corresponding to the peak of the histogram (images n^o 41, 132, and 100).

However, among the histograms of votes, there were as well numerous cases for which no peak stood out, making the selection of a representative category less obvious with the basic approach, as it can be seen in the histograms in Figure 8.3.

At this point, the representation using the histogram of votes became less meaningful, since the categories that were closely related in significance, and therefore close to each other in the three-dimensional schema of the dataset in Figure A.2, became scattered in the histogram representation, due to the linearity of a diagram of this type. Thus, what in Figure 8.3 seem to be three representations of multiple dissociated peaks actually shows the hesitations of the participants between categories which do not differ among themselves very strongly.

Because of this ambiguity in the interpretation of the histograms of votes, a new solution was needed for determining the representative category for each image by also taking into account the various relations between the 18 categories.

As a consequence, a specific metric inspired from the formula of inertia was designed, which would calculate the overall relevance of each voted category in function of the heights of the bars in the histograms of votes, but also in function of how the categories corresponding to these bars were related in the three-dimensional structure defining the dataset. By using this metric

for a given image, an *inertia* value could be calculated for each voted category, leading to a new histogram for that image, called *histogram of inertias*.

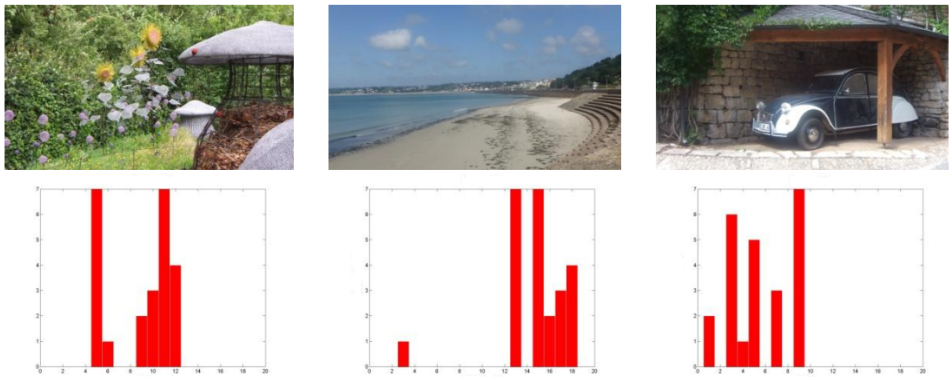


Figure 8.3: Examples of images that cannot be easily classified in one of the 18 categories – images n^o 29, 61, and 73.

The *inertia metric* used is:

$$inertia\,(category) = \sum_{j \neq i} (number\ of\ votes\,(category\ j)) \cdot (distance\ (category\ i,\ category\ j))^2, \quad (8.1)$$

where the *distance* matrix was calculated using an adaptation of the chessboard distance from the two-dimensional space to the three-dimensional space, to make it suitable to our three-dimensional stereoscopic dataset structure explained in Figure A.1.

Applying this formula to a given voted category led to a *high inertia* value when that category was little or not at all related to the other categories with many votes and to a *low inertia* value if that category was strongly related to the other voted categories. The case of a small value for a category thus indicated that there was a consensus between the participants on the majority of the characteristics of that image and that the choice of that category was not the result of random voting or of opinions that were too disparate from the general trend.

Consequently, this more elaborated method for detecting the representative category of an image was focused on the analysis of the histograms of inertias. A category could be declared representative for the image if the highest peak or one of the two highest peaks of the histogram of votes corresponded to the smallest positive value in the histogram of inertias. This would mean that the category that gathered the maximum number of votes was also the one that concentrated the general trend of all the votes for that image. Otherwise, when the maximum of votes did not correspond to the minimum inertia, the image would be declared as *impossible to classify*.

Figure 8.4 explains this logic with the example of image n^o 49. For a more intuitive illustration, in the histograms of inertias, the red bars always indicated the category or the categories

corresponding to the peak(s) of the histogram of votes. Thus, it can be noticed in Figure 8.4 that the category corresponding to the peak of the histogram of votes (red histogram) corresponded as well to the minimum value of the histogram of inertias (blue histogram), indicating without ambiguity that this was the representative category for image n^o 49.

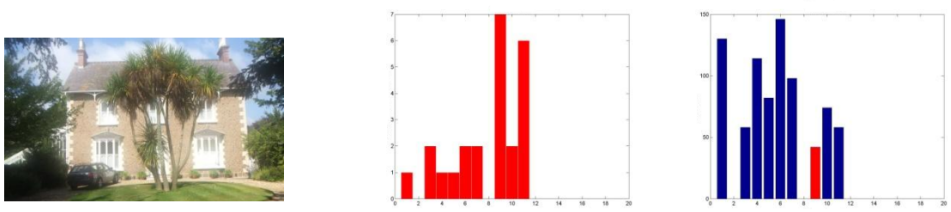


Figure 8.4: Example of the identification of the representative category of an image using the histogram of inertias – illustrated for image n^o 49.

Following these considerations, the three ambiguous situations previously presented in Figure 8.3 became easier to interpret, as Figure 8.5 shows. For each of the first two images (n^o 29 and n^o 61), the solution was given by the category with the smallest inertia value among the two corresponding to the two maximum peaks of the histogram of votes. However, for the third image (n^o 73), the peak of the histogram of votes did not correspond to the minimum value in the histogram of inertias, therefore the classification could not be successful.

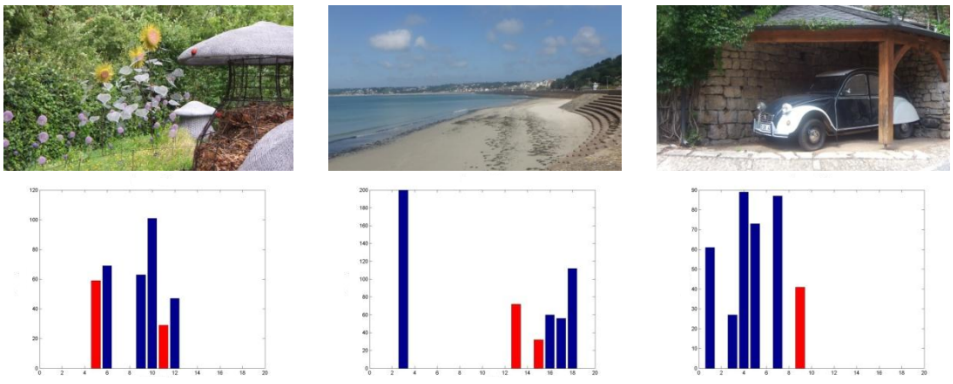


Figure 8.5: Eliminating the ambiguity in the selection of the representative categories of the images in Figure 8.3.

After the analysis of the data using this *inertia method*, a satisfying rate of 77 % was achieved for the classification, with 121 images well classified in one of the 18 categories and 37 images for which the classification was not possible.

Despite the fact that consensus was achieved in judging the images according to the categories defined at the beginning of the experiment, an unexpected result showed up, that can be observed in Figure 8.6: the distribution of the images in the 18 categories was not at all balanced, several categories being even empty (represented in red). Moreover, it could be observed that all the empty cases were representing categories with the condition monocular depth cues absent. Also, the rest of the cases corresponding to the same condition contained very few images (represented in blue).

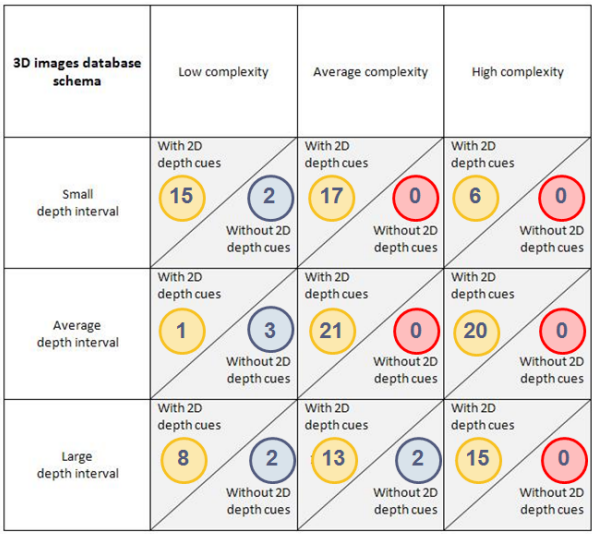


Figure 8.6: The distribution of the correctly classified 121 images in the 18 categories, using the inertia method.

There were several explanations possible for these results. The fact that the participants considered that there were not many images without monocular depth cues in our dataset could have been determined by the imperfect choice of the dataset at capture, by the imprecise description of the *monocular depth cues* characteristic to the participants, by the incorrect or different comprehension of this characteristic by them, or just by the difficulty for the photographers of finding situations in the real world where no monocular depth cues were present.

The unbalanced distribution of the images led to the necessity of a new logic in the processing of the data and this is illustrated by Figure 8.7. The monocular depth cues characteristic was suppressed from the dataset structure, since no subsequent analysis could lead to useful results, given the lack of sufficient images on certain positions of the original dataset structure. The two other characteristics – the complexity and the depth interval – were maintained.

8.2.2.1 Adaptation of the data processing approach

After the abandon of one of the three criteria of classification, for clarity in the following analyses and for improving the classification rate, another approach seemed more interesting: the classification of the images in function of only one of the two valid criteria at a time.

This led to only 6 categories of images, 3 categories derived from the classification only in function of the *complexity* and 3 categories obtained by classifying the images only in function of the *depth interval*.

The translation of the initial 18 categories to the 6 new ones has been done in the following manner:

- Category 1 – *low complexity* – old categories: 1, 2, 7, 8, 13, 14.
- Category 2 – *average complexity* – old categories: 3, 4, 9, 10, 15, 16.

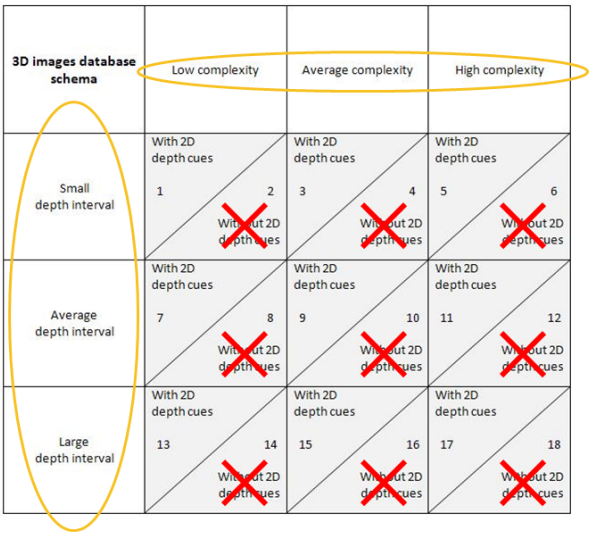


Figure 8.7: The new dataset structure, after suppressing the *monocular depth cues* dimension. The only dimensions that remain valid are *complexity* and *depth interval*.

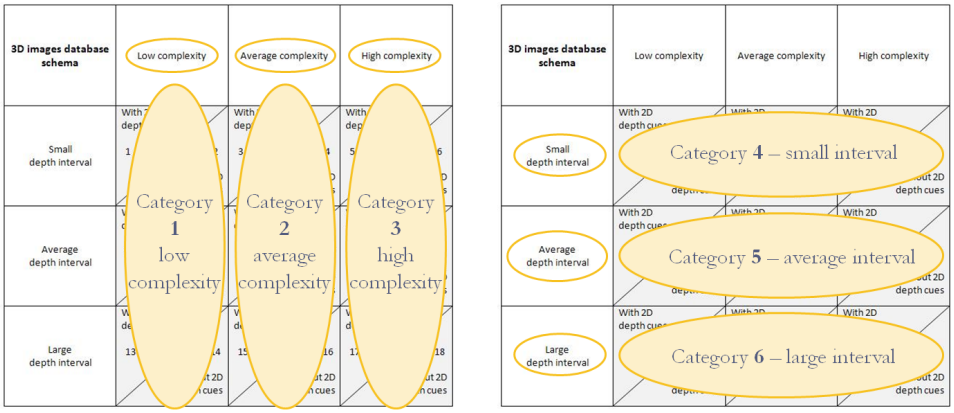


Figure 8.8: The new dataset classification in function of a unique criterion at a time.

- Category 3 – *high complexity* – old categories: 5, 6, 11, 12, 17, 18.
- Category 4 – *small depth interval* – old categories: 1, 2, 3, 4, 5, 6.
- Category 5 – *average depth interval* – old categories: 7, 8, 9, 10, 11, 12.
- Category 6 – *large depth interval* – old categories: 13, 14, 15, 16, 17, 18.

The number of votes corresponding to each of these 6 new categories was considered to be the sum of the votes corresponding to each of the old categories that they included.

For determining the two representative categories for each image (one category relative to the complexity and one category relative to the depth interval), histograms of votes have been used, in a similar fashion as at the beginning of this analysis. The initial table containing the

raw classification results allowed the re-processing of the data for the creation of two histograms per image, one for each of the two remaining characteristics considered.

Since each histogram only had three bars corresponding to the three possible conditions of each criterion, the representative category was easier to be chosen as the category corresponding to the peak of the histogram. In the few cases where the histogram had two maximum peaks, the image was declared *impossible to classify*.

Following this adapted approach for the classification, in function of only one criterion at once, very high rates of classification have been achieved: 97 % (with 154 images well classified) for the classification of the database in function of the complexity criterion and 98 % (with 155 images well classified) in the case of the depth interval criterion. The distribution of the images in the 6 new categories was also very satisfying, proving good balance, as Figures 8.9 and 8.10 or Table 8.1 show.

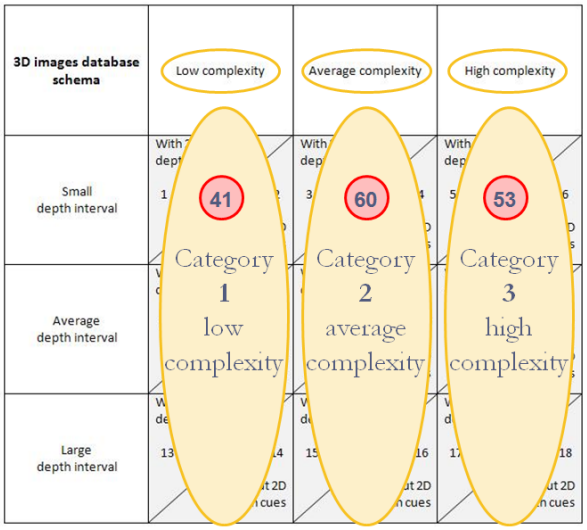


Figure 8.9: The distribution of images per category, for the classification in function of the *complexity* criterion.

Table 8.1: The final distribution of all the dataset images per category.

Complexity			Depth interval		
Low	Average	High	Small	Average	Large
41 images	60 images	53 images	50 images	52 images	53 images

8.2.3 Conclusions after Part I

The dataset classification thus obtained constituted a solid basis for the selection of the images to be used during the second part of our study and also for the analyses to be made on its results. In this manner, it allowed later the exploration of various correlations between the subjective

appreciations on the stereoscopic images and their two considered features, *complexity* and *depth interval*.

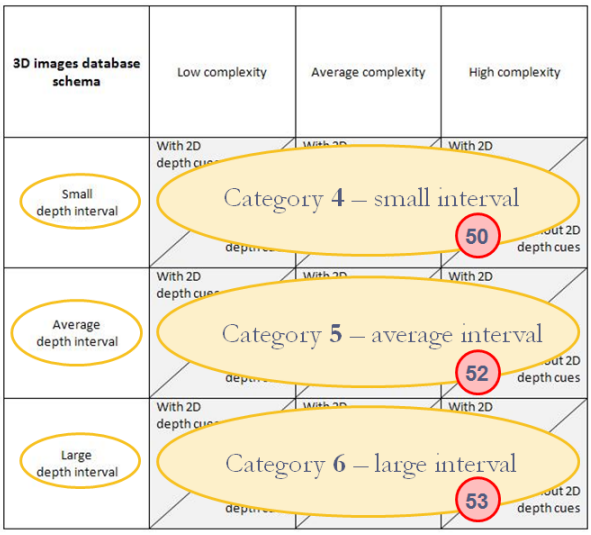


Figure 8.10: The distribution of images per category, for the classification in function of the *depth interval* criterion.

8.3 Selection of the stimuli set for Part II

Starting from the structured database that we obtained after the first part of our qualitative experiment, we were able to select, in a precise manner, 24 images as the test material for the second part of our study. In order for this selection to lead to a dataset as controlled as possible, we considered three criteria to model our choices.

The first criterion of selection was to *cover all the 6 categories* previously defined, by taking into account the subjective classification made by the participants. We considered that a minimum of 4 images per category was sufficient.

The second criterion of selection was related to the depths represented on screen, which are given by the disparities that were present in the stereoscopic pairs. A stereo-matching algorithm implementing the SIFT[1] feature detection method was used for computing the disparities of a large number of feature points in each of the 24 images (more details in Appendix E.2). This allowed us to determine the minimum and the maximum disparities contained by each of the stereoscopic images, *i.e.* their disparity ranges, and then to *select images of varying represented depths* for Part II of our study.

And the third criterion of selection was to *cover the largest range of semantic content types possible*.

Consequently, by following these three sets of guidelines, we managed to obtain the published *MOOV3D – GIPSA Stereoscopic Dataset 1*, with its precise structure.

[1] scale-invariant feature transform

The distribution of the selected 24 images according to the complexity and depth interval criteria is given in Table 8.2 and the categories to which the images belong have already been shown in Figure A.1.

Table 8.2: The distribution of the 24 dataset images per category.

Complexity			Depth interval		
Low	Average	High	Small	Average	Large
4 images	8 images	12 images	9 images	8 images	6 images

The graphical representation of the horizontal disparity ranges of each of the 24 images, given in pixels for the original-resolution images of 3584×2016 pixels, is in Figure 8.11.

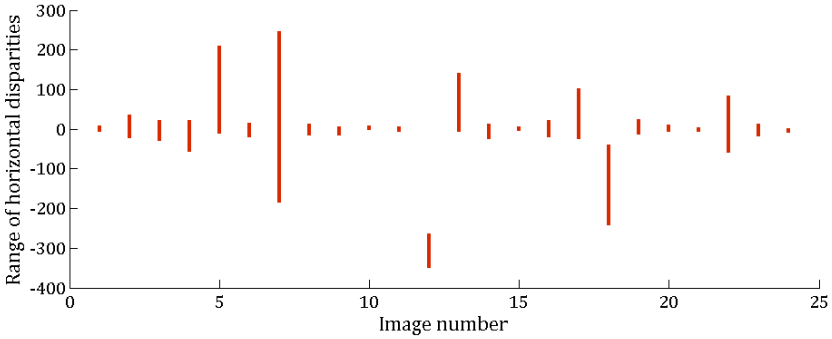


Figure 8.11: The horizontal disparity ranges of the 24 images.

And the semantic diversity of the images chosen can also be observed in Table A.1.

Part II of the exploratory study was exclusively based on this collection of 24 stereoscopic images.

8.4 Part II – Visualization and verbalization

8.4.1 Data collected

Since the second part of our exploratory test consisted in participants watching stereoscopic images and making comments on the perceived quality, the raw data stored during the test were represented by the *recordings* of the observations made by the participants.

The preparation of the collected data for the subsequent processing was hence the extraction of the text from the audio recordings and the organization of this text in an *observations table* corresponding to all the images and to all the participants (also called *judges* in this context, since they made judgments on quality). This led to the structure of size 24×24 that is schematized in Table 8.12, with each case representing all the observations made by one participant on one image.

image / judge	judge 1	...	judge i	...	judge 24
image 1		...		...	
...	...	...	...	...	...
image j		...		...	
...	...	...	...	...	...
image 24		...		...	

Figure 8.12: The organization of the experiment data in the observations table – transition from audio to text and from brute text to a structured text table.

8.4.2 Data processing and results

The analysis of the observations table started with the scanning of all the text in search of the *key concepts* used by the participants. The *quality* and the *comfort* characteristics stood out during this text processing, as expected. But it turned out that the participants had also shown much importance to the idea of *realism* and to a particular artifact, the *cardboard effect*, and had made detailed observations related to them.

These four characteristics were subsequently studied in detail by extracting *key phrases* referring to each of them from the text table. Based on these key phrases, *numeric ratings* could be associated to each case of the text table for each of the characteristic, by considering adapted scales, as Figure 8.13 shows. The choice of the complexity and type of the *scale* for each characteristic was determined by the degree of detail that was present in the comments made by the participants or by the nature of the characteristic itself.

Five-rating scales were chosen to represent the data for quality and comfort, a three-rating scale was used for realism, and a binary scale was considered for the cardboard effect. The tags associated to each of these ratings are the following:

- for *quality*:
  - very low quality (-2),
  - low quality (-1),
  - neutral (0),
  - high quality (+1),
  - very high quality (+2),
- for *comfort*:

  - very uncomfortable (-2),
  - uncomfortable (-1),
  - neutral (0),
  - comfortable (+1),
  - very comfortable (+2),

· for *realism*:

  - artificial (-1),
  - neutral (0),
  - realist (+1),

· for *cardboard effect*:

  - cardboard effect present (-1),
  - cardboard effect absent (0).

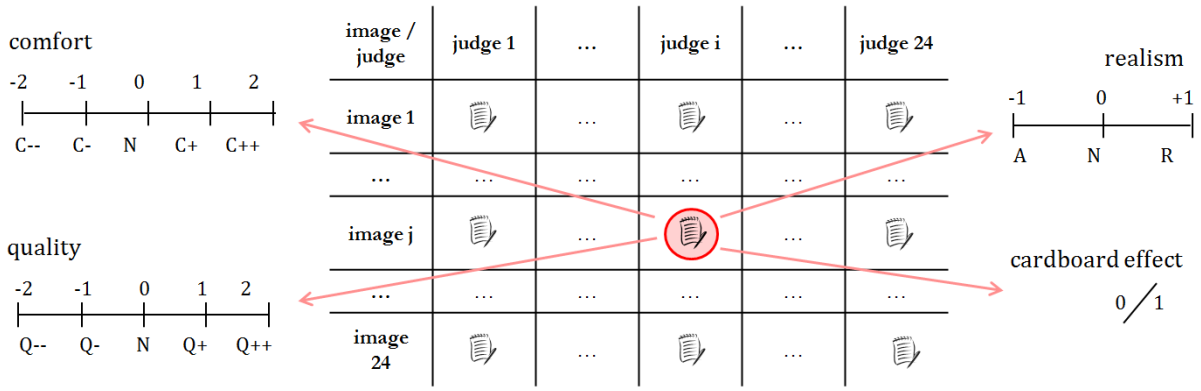


Figure 8.13: The ratings used in the text analysis.

In order to illustrate how the association between key words and ratings has been made, in Table 8.3 there are a few examples of such associations, where each rating is in brackets, after the group of words to which it was attributed.

All the subsequent studies were made by analyzing these extracted ratings. The method used for computing an overall rating per image from the quality and comfort ratings was *scales aggregation*, in the form of the median value of all the votes for that image, as explained in Appendix B.8. For realism, we used however the adapted formula:

$$realism\ rating\ per\ image = (noP - noN)^2 \cdot (noP - noN)/(noP + noN), \tag{8.2}$$

where noP and noN represent the number of positive votes and the number of negative votes for that image. And, in the case of the cardboard effect, the rating per image was given directly by the number of participants that made comments related to the presence of the artifact in that image (situations marked with *1* in the numeric table). The aggregated ratings thus obtained can be observed in Figure 8.14.

Table 8.3: Examples of extracted key words and their associated ratings.

Quality	Comfort	Realism	Cardboard effect
very bad quality (-2)	*really annoying, uncomfortable* (-2)	*it's super artificial* (-1)	*impression of successive layers* (1)
the quality is below average (-1)	*a little bit annoying* (-1)	*it's not like in reality* (-1)	*no continuum in depth* (1)
average quality (0)	*I don't feel annoyed, but it's not comfortable either* (0)	*like in real life* (1)	*as if each object was flat* (1)
it's rather a good quality image (+1)	*generally speaking, it's comfortable* (1)	*natural* (1)	*a little like made of cardboard* (1)
very, very good quality image (+2)	*very relaxing* (2)	*I see myself projected in the image* (1)	*rather like having three planes, than a volume* (1)

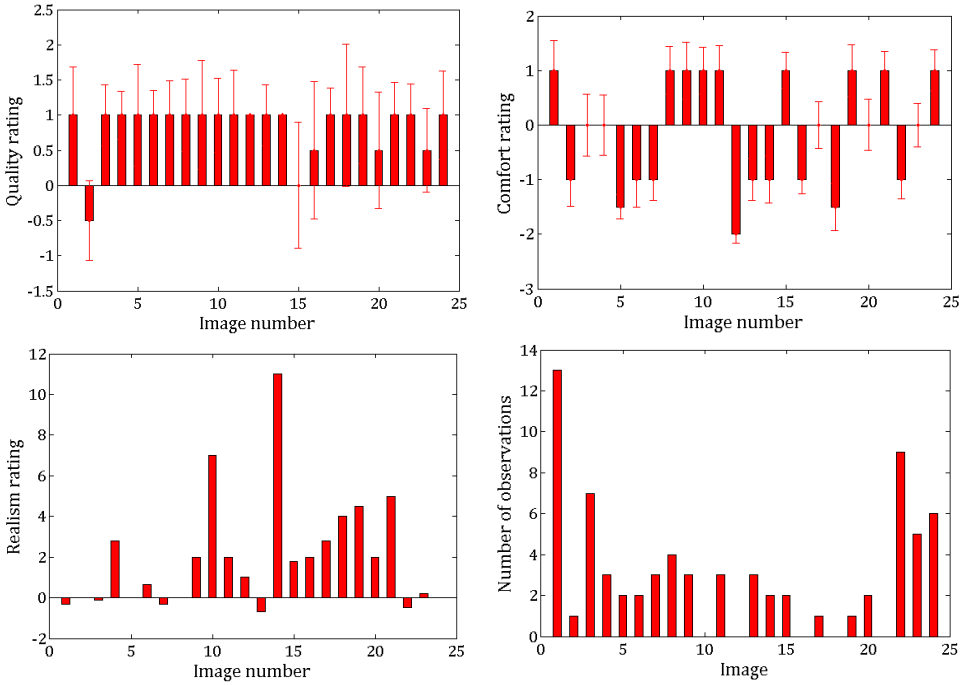


Figure 8.14: The subjective ratings per image obtained from the ratings table.

8.4.2.1 Results on quality

The fact that almost all the stereoscopic images were judged as of high quality by the participants, as it can be seen in Figure 8.14 (a), proved that the notion of quality was an ambiguous term during the experiment. The participants have been asked to express their opinions on the "quality of the stereoscopic images" and the prior expectations were that they would include in this notion the totality of the aspects that contribute to the overall stereoscopic experience.

On the contrary, the responses provided by the participants on the notion of quality were

proven to refer mostly to the 2D quality of the views composing the stereoscopic image. This behavior can be explained by the fact that the 2D quality was very similar throughout the database, since all the images have been taken with the same camera in the same conditions, and the only images that obtained a lower quality score were those that contained some kind of 2D degradation (motion blur or elements that were perceived as lack of contrast).

Our conclusion is that the concept of *quality* should always be defined during a subjective experiment in function of the context. In our case, where we explore the perceptions triggered by stereoscopic content, we consider that *image quality* can successfully constitute one of the dimensions of the complex notion of *stereoscopic image quality* and cannot be substituted to it. Moreover, we consider that the *image quality* concept should be related exclusively to the *quality of the 2D views* that make up the stereoscopic data.

8.4.2.2 Results on comfort

The subjective ratings on comfort were first analyzed in relation to the *complexity* categorization of the 24 images. We could observe, as illustrated in Figure 8.15, that, when grouping the ratings on comfort in function of the three levels of complexity of the images to which they correspond, more images rated as uncomfortable were in the high complexity group and only one image judged as uncomfortable was in the low complexity group. This indicated a possible influence of complexity on comfort. However, we also performed a statistical analysis using the one-factor Anova test and the multiple comparison test, explained in Appendix B.6, and the results showed that there were no significant differences among the complexity categories for our comfort ratings ($p = 0.4276$). Thus, the influence of complexity on comfort could not be validated in the context of our experiment.

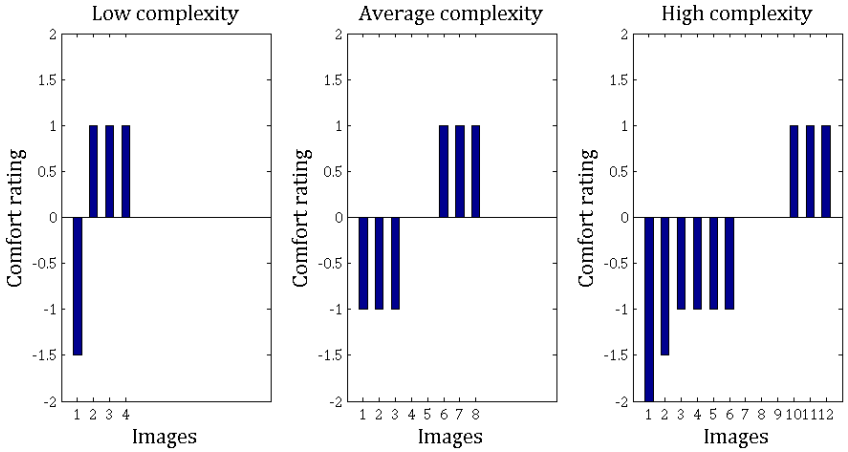


Figure 8.15: The subjective comfort ratings grouped in function of the complexity of the corresponding images.

When grouping the 24 comfort ratings in function of the *depth intervals* of the corresponding images, one category stood out as significantly different compared to the others. The category of images with large depth intervals contained only images with positive comfort ratings, while in the other two categories there were mostly images with negative comfort ratings, as Figure 8.16 shows. The fact that large depth intervals were related to positive comfort in the context of the assessment of our database was confirmed by the Anova test ($p = 0.0074$) and by the

multiple comparison test. This result showed that the comfort perceived can depend on the distribution of the real depths in the photographed scene, or on the capture configuration that this distribution of real depths might impose.

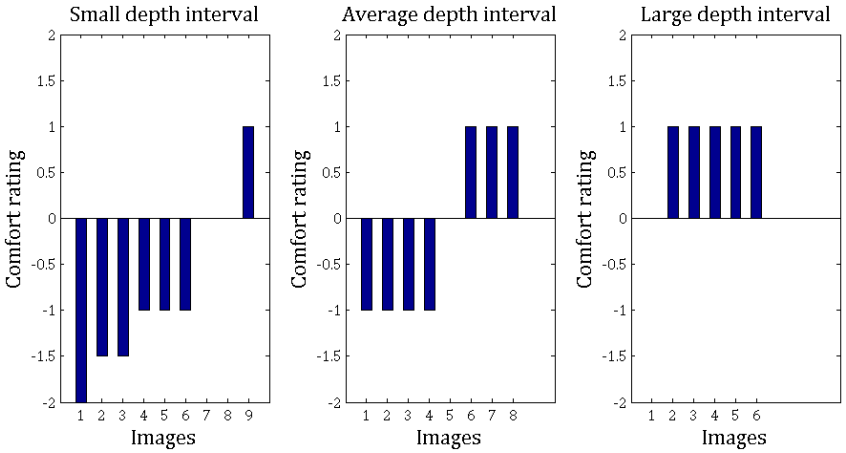


Figure 8.16: The subjective comfort ratings grouped in function of the depth interval of the corresponding images.

A third test was done in order to verify the relation between the comfort ratings and the *horizontal binocular disparities* in the 24 images. The correlation index of -0.5716 was sufficiently significant to indicate an inverse correlation between the two. The graphical representation in Figure 8.17, where the results per image are ordered in function of the comfort ratings, also shows that, for small disparity ranges, the comfort was usually positive, while, for large disparity ranges, the comfort ratings were always negative in our experiment. Thus, the results obtained indicated a negative influence of large disparity ranges on comfort and these results are consistent with our expectations, confirming that images with large represented depths on screen need more effort for the fusion of their content, thus implying more visual discomfort.

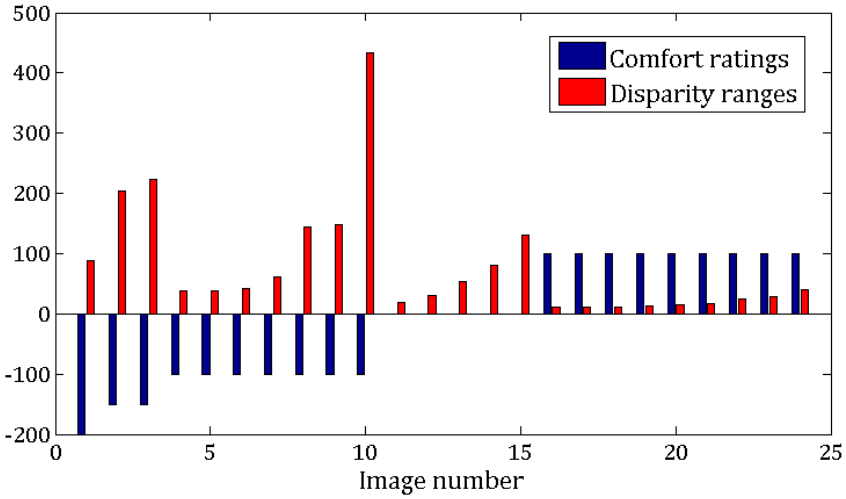


Figure 8.17: The comfort ratings and the disparity ranges per image, ordered in function of the comfort.

Following the numerical analysis of the data on *comfort*, our conclusion is that this perceptual

attribute is mostly influenced by the *horizontal disparities* that are present in the stereoscopic images, but also by the *original depths* in the real scene that was photographed. Other characteristics of the stereoscopic content can have an influence on comfort as well, but further more precise studies are necessary for an in-depth view on this subject.

8.4.2.3 Results on realism

For judging the realism percept in relation to the *complexity* of the stereoscopic images, the 24 realism ratings were classified in function of the three complexity categories, as presented in Figure 8.18. The median values of the three groups of ratings were determined as distinct and the results of the statistical tests showed significant differences among them ($p = 0.0043$). Thus, the compositional complexity appears to be influencing realism in the sense that images of low complexity have been considered more realistic and images of high complexity have been considered more artificial during our visualization test.

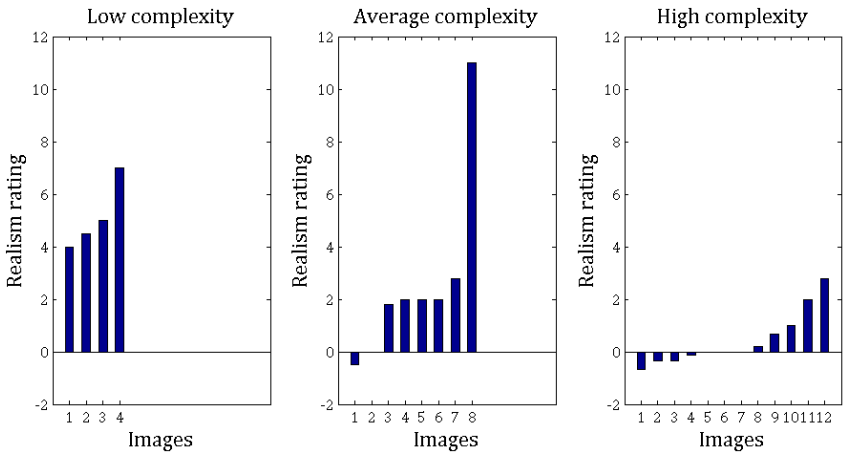


Figure 8.18: The realism ratings grouped in function of the complexity of the corresponding images.

In the case of the real *depth intervals*, neither the graphical representation nor the statistical results ($p = 0.8739$) could give indications on a correlation between this characteristic and the realism perceived.

The correlation index computed for the realism and the *disparity range data* (0.0706) was not relevant either, suggesting that the disparities in the stereoscopic images did not influence the realism percept induced. Our test hypothesis, which supposed that the realism could be influenced by the presence of negative disparities that give a sensation of immersion, was invalidated.

In the mean time, a correlation index of -0.4689 between the realism ratings and the *cardboard effect* ratings suggested that the presence of the cardboard effect made the stereoscopic images seem more artificial. This conclusion was equally illustrated by the graphical representation of realism and cardboard effect in parallel, in Figure 8.19, where it can be observed that for the images where the cardboard effect was pronounced, the ratings on realism were negative or neutral, suggesting a sensation of artificiality or no percept of realism.

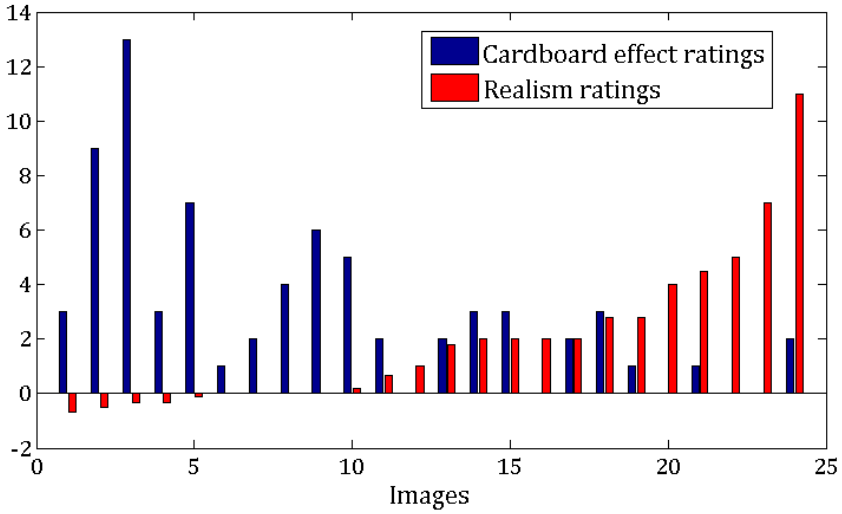


Figure 8.19: The cardboard effect ratings and the realism ratings per image, ordered in function of the realism.

The observations that we could make related to the *perceived realism* of the stereoscopic images encouraged us to conclude that among the tested factors, the *cardboard effect* had the strongest negative influence. The fact that *high complexity* images also triggered a sensation of artificiality for the images of our database could also be associated to the cardboard effect, which was better perceived in images with a larger number of elements.

Since the cardboard effect is a geometric distortion, we believe that other *geometrical distortions* could influence as well the percept of *realism*, but this hypothesis needs to be submitted to test. We will however refer to the general category of physical parameters which trigger the *cardboard effect* as *3D geometry parameters*.

8.4.2.4 Overall results

Since one of the purposes of the work presented in this manuscript was to identify perceptual attributes that could lie at the basis of a *stereoscopic image quality model*, as defined in the Stereoscopic Image Quality Circle (SIQC), we tested the interactions between the three perceptual attributes that we could identify during this exploratory study: *quality* (that proved to be actually a *2D image quality* attribute), *comfort*, and *realism*. The correlation coefficients calculated for the subjective ratings on quality and comfort, quality and realism, and comfort and realism (0.0451, 0.1364, and 0.0931) showed a lack of dependency between these three attributes. This was a sufficient indicator of the fact that all of these three perceptual attributes could be included in an independent manner in the elaboration of a model that evaluates the *perceived stereoscopic image quality* (also referred to during our experiments as the *overall stereoscopic quality perceived*).

As a consequence, the first *stereoscopic image quality model* that we could propose after this exploratory test is shown in Figure 8.20. This model is completely compatible to be integrated in the Stereoscopic Image Quality Circle structure, in the sense that it is a model that links the *observer ratings* on the *overall percept* to a set of *perceptual attributes* that lie at the basis of the

overall judgment on quality. Moreover, this model illustrates the relations that exist between the perceptual attributes that we identified and a set of *physical stereoscopic image parameters*.

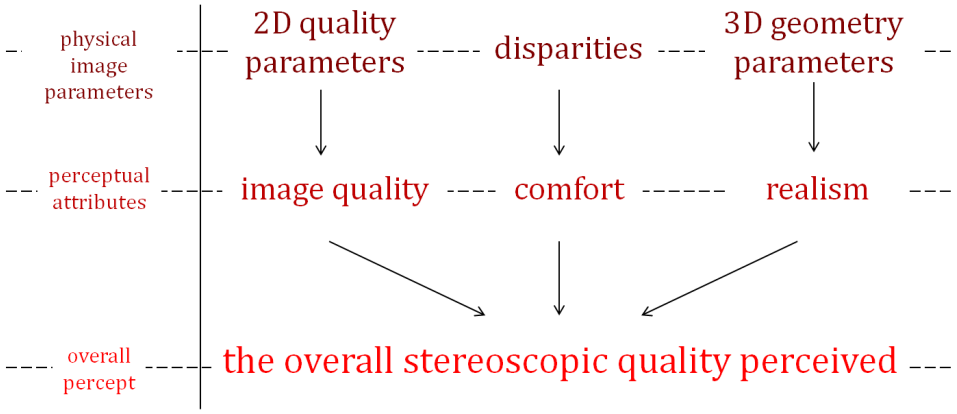


Figure 8.20: The proposed stereoscopic image quality model.

Conclusions

The purpose of our qualitative study was to determine which perceptual attributes could lay at the basis of a stereoscopic image quality model. Our exploratory test proved to be rich in results. Among its conclusions, we consider of highest importance the fact that, starting from a database of varied and controlled content, we could extract three different perceptual attributes – *image quality*, *comfort*, and *realism* – that were demonstrated to influence the *overall perceived quality of stereoscopic images*.

Since we could identify the *2D quality impairments*, the *levels of horizontal disparities*, and the *3D geometry distortions that produce the cardboard effect* as physical properties of the stereoscopic data that influence the three perceptual attributes mentioned, we consider our results as *a starting point for the elaboration of a stereoscopic image quality model* to be integrated in the Stereoscopic Image Quality Circle proposed in Section 5.2.1.

The fact that the *depth* element is missing from our proposed model, compared to other models in the literature [Seuntiens 2006a, Lambooij 2011, Chen 2012b], can be naturally justified by the fact that depth is a dimension situated at the level of the physical properties in the SIQC and not at the perceptual attributes level, since for given capture and visualization conditions, the depth perceived for a certain point is a measurable value. Therefore, even if the depth is visible to the observers, we consider that they do not evaluate it directly, but that its appreciation is indirectly reflected by the *comfort* and *realism* perceptual attributes.

Chapter 9

A Quantitative Study

Introduction

Following the results of our exploratory study, we were able to draw a first sketch of a stereoscopic image quality model, illustrated in Figure 8.20. However, in order for this model to be *validated, then refined*, more precise investigations were necessary. As Chapter 4 has shown, the suitable tests in this context are the *quantitative explorations*, in which previously known perceptual attributes are rated by human participants.

The current chapter shows therefore *our quantitative approach* which consisted in performing two experiments in order to test the validity of our first model and to quantify the relations between its components.

The following sections give comprehensive information on these two experiments, on the data analysis process and on the results obtained after each one of them.

9.1 Experiment 2

The first quantitative experiment that we implemented can be seen as a direct sequel of the qualitative study presented in the previous chapter. It searched to assess, by quantitative means, the human judgments on the three perceptual attributes identified (*image quality*, *comfort*, and *realism*) and also on the *overall percept of stereoscopic quality*, given a dataset of stimuli precisely built for this purpose.

The questions that this experiment tried to answer with precise subjective measures are synthesized in Figure 9.1.

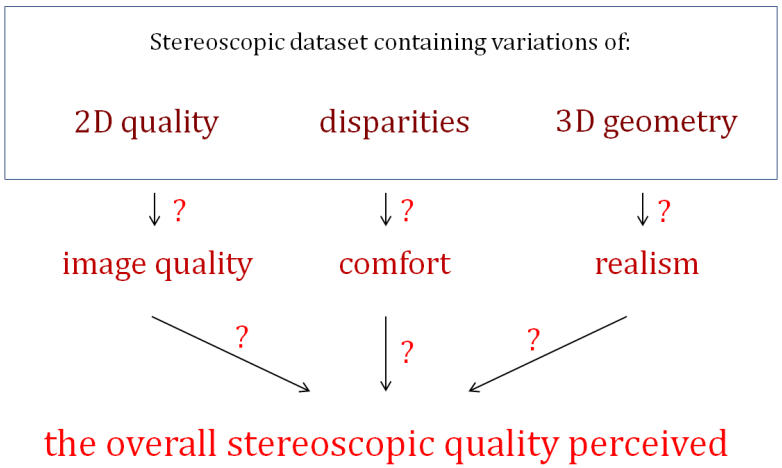


Figure 9.1: The objectives of Experiment 2.

9.1.1 The choice of the stimuli

Given the problem raised by Figure 9.1, we created the *MOOV3D - GIPSA Stereoscopic Dataset 2*, described in Appendix A.2, as the stimuli set to be evaluated by the participants during this first quantitative experiment.

This stimuli set thus contains the variations of blur, of crossed disparities, and of camera-object distance at capture that are detailed in Table A.3. This structure was justified by the purpose of this quantitative experiment, in the sense that, in order to be able to understand the way the participants judge the variations of image quality, of comfort, and of realism and how they integrate them into a global percept of stereoscopic image quality, the dataset needed to

contain images that would trigger differences in the perception of all these attributes.

In this manner, the variations of *blur* that we introduced in our stimuli set were intended to induce variations in the *image quality* perceived; the various *crossed disparities* were intended to influence the *comfort* perceived; and the different *capture distances* were intended to produce various distortions of 3D geometry and thus to have an impact on the impression of *realism*. These three types of variations are detailed in the following paragraphs.

Blur, the first variable considered, was subjectively added by us, identically on the two views of the stereoscopic pairs, such that its presence would be visible, but not strongly annoying.

The levels of *crossed disparities* were chosen by considering the observations on the maximum disparities supported by the human visual system in [Pastoor 1995]. We thus placed some of our stimuli at the limits of the comfortable stereoscopic visualization from the point of view of the represented depths.

Regarding the *capture distances*, they were chosen as a dataset variable by having as a starting point the representation in Figure 6.8, where a stronger compression of the perceived depths can be observed with our system for objects photographed from further away. Since the exploratory study that we performed showed an influence of the presence of the cardboard effect on the feeling of realism, we computed this effect for our model in function of the camera-object distance, inspired by what in Figure 6.8 seemed to resemble precisely to a pronounced cardboard effect for objects photographed from large distances. The formula used is explained in Appendix E.3. We thus observed the evolution displayed in Figure 9.2, showing indeed a stronger *cardboard effect* for objects photographed from further away. This led us to creating the MOOV3D – GIPSA Stereoscopic Dataset 2 with variations of capture distance, with the expectancy to generate variations in the percept of realism as a consequence.

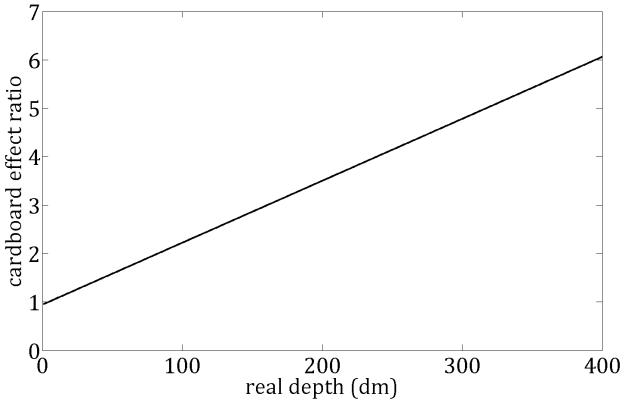


Figure 9.2: The evolution of the cardboard effect with the capture distance.

Dataset 2 was thus considered by us to be adapted to the purpose of our first quantitative experiment, which allowed us first to test our hypothesis of dependence between the considered physical stereoscopic image parameters and the perceptual attributes of our model, then to proceed to the refinement of our proposed stereoscopic image quality model.

9.1.2 Experimental protocol and participants

For the implementation of this first experiment, we followed the *quantitative protocol* described in Section 7.3.

The same procedure was reproduced both on the 3D TV screen and on the prototype 3D glasses illustrated in Section 6.3. For both displays, the participants were supposed to rate the four main concepts from our first proposed model: *image quality*, *comfort*, *rendering realism*, and *overall stereoscopic quality*, all of them explained in detail on the task sheets that they received before the experiment, which are reproduced in Appendix D.2.

Since we implemented the same experiment both on a 3D TV screen and on 3D glasses, we considered including an extra-task, during which we would ask the participants to also assess the level of *immersion* that the stereoscopic images displayed induced. This task would have had the purpose of also evaluating a dimension that we considered fundamentally different between the two displays. Its absence from the original model proposed after our exploratory study was justified by the fact that this study had been performed exclusively on the 3D TV screen. Therefore, immersion did not stand out during this exploratory study as a representative attribute for the 3D TV visualization, which was in accordance with our expectation of lower immersion for the 3D TV. However, despite this logic, we were constrained to only implement the evaluation of the initial four concepts, given the already dense structure of the experiment, implying a large number of participants and a very long duration.

Therefore, Experiment 2 was actually fragmented in 8 individual sub-experiments, each of them identified by a task for the participants and a display technology. This structure is schematized in Figure 9.3.

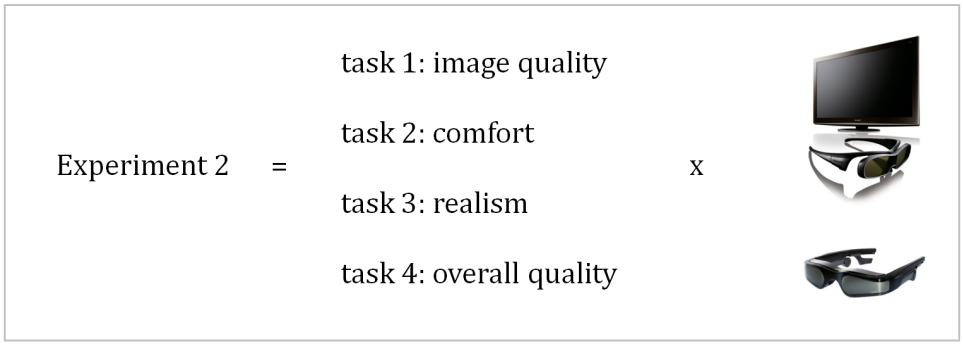


Figure 9.3: The structure of Experiment 2.

After discarding all the candidate participants that did not pass the vision tests described in Section 7.1.1 (34 participants discarded in all), a total of 102 valid candidates participated to this test. 56 persons took part in the experiments on the 3D TV screen and 46 persons in the experiments on the prototype 3D glasses. They were distributed per task as shown in Table 9.1.

In this population sample, 37 persons were female and 65 were male. The age range was 18 to 57, with an age mean of 25 and an age median of 23.

The exact test plan of this quantitative experiment is included in Appendix D.2.

Table 9.1: The distribution of participants per task and per display during Experiment 2.

	tasks			
	image quality	comfort	realism	overall quality
3D TV screen	15	15	15	11
3D glasses	13	12	11	10

9.1.3 Data collected

The Soft Eye application introduced in Section 7.1.6, which was used for the stimuli presentation during the experiment, also performed the automatic storage of the data collected. These data were provided at the end of the participation of each person in the form of a text file which stored a large number of information, like the identification names of the stereoscopic images, the order in which these images were displayed, the ratings the participant gave to each of them, or the time benchmarks.

At the end of the experiment, a collection of 102 text files, one per participant, was available for further processing.

9.1.4 Data processing

9.1.4.1 The data extraction

First, the test files have been automatically scanned by a Matlab code in order to extract all the ratings given during the quantitative study in 8 matrices, one for each individual experiment. The size of such a results matrix was $N \times TC1 \times TC2 \times TC3 \times K \times R$, where the notations stand for:

N: the number of participants;

$TC1$: the number of test conditions for the first variable (blur);

$TC2$: the number of test conditions for the second variable (crossed disparities);

$TC3$: the number of test conditions for the third variable (capture distance);

K: the number of test images;

R: the number of repetitions of the test stimuli display.

Given the explanations on the dataset used (in Appendix A.2) and on the quantitative protocol implemented (in Section 7.3), the actual values of the enumerated parameters for each of the 8 tests making up Experiment 1 were:

N: different for each of the 8 tests, can be read in Table 9.1;

$TC1$: 3;

$TC2$: 3;

$TC3$: 3;

K: 4;

R: 2.

Therefore, the size of each of the 8 ratings matrices of Experiment 2 was actually $N \times 3 \times 3 \times 3 \times 4 \times 2$.

9.1.4.2 The distribution of the scores

Next, we proceeded to analyzing the distribution of the participant scores across all the $TC1 \times TC2 \times TC3 \times K$ test stimuli. We only show here the results concerning the second repetitions of the experiments, since, as already explained in Section 7.3, we only considered the second repetition for computing the final results.

Consequently, we created the histograms of scores for each of the 8 experiments, which show, for each possible score, the number of times that the score was attributed to a stimuli. Then, we analyzed the *kurtosis* measure corresponding to each one of the histograms. The kurtosis value can be seen as a descriptor of the shape of a given distribution and more details on its computation can be found in Appendix B.2. The 8 histograms and their corresponding kurtosis coefficients are shown in Figure 9.4 and Table 9.2.

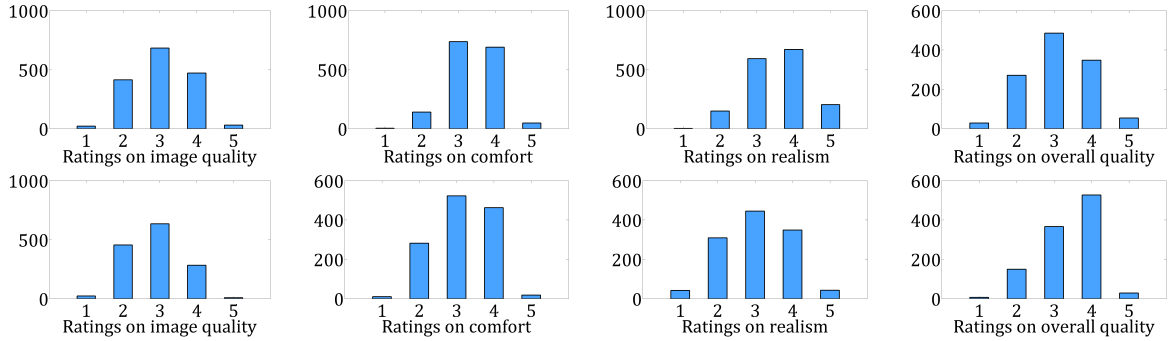


Figure 9.4: The histograms showing the scores distribution for each of the 8 tests of Experiment 2 (the top line shows the scores of the tests on the 3D TV and the bottom line the scores for the 3D glasses).

The fact that all the kurtosis values belong to the $[2, 4]$ interval indicates that the score values of Experiment 2 can be considered *normally distributed* [ITU 2012b] and that *parametric statistical methods* could be later used for processing these data.

Moreover, the fact that the kurtosis coefficients are closer to the inferior limit of the $[2, 4]$ interval indicates that the variance of the votes is not due to a concentration of the votes exclusively at the center of the scale, but that the tails of the distribution also concentrate an important number of votes, configuration referred to as *platykurtic*. This is confirmed by the

appearance of the histograms in Figure 9.4 and shows that the entire scale has been used by our participants.

Table 9.2: The kurtosis measures of the score distribution per task and per display during Experiment 2.

	image quality	comfort	realism	overall quality
3D TV screen	2.1883	2.5471	2.5572	2.4415
3D glasses	2.3770	2.3515	2.4458	2.5433

9.1.4.3 Data normalization

In order to compensate for the variations in the way the participants used the same rating scale, another data processing step was necessary before going further: the *normalization* of all the values in order to adjust them to a common scale. We chose to do the normalization by computing the *standard scores* corresponding to each of the original scores, then by remapping the standard scores obtained to the original interval of our scale, *i.e.* $[1, 5]$. The normalization formula applied can be found in Appendix B.5.

9.1.4.4 Participant screening

A last step before proceeding to the computation of the results was to check the distributions of the ratings per participant and to remove the data corresponding to participants with anomalous scores.

For this, we used two approaches. First, we applied the *ITU screening method* [ITU 2012b]. This method verifies if there is any participant whose ratings differ too much from the general trend of the ratings of the entire experiment. Any such participant was consequently discarded. The mathematical details on the logic behind this screening method can be found in Appendix B.3.

Second, we also checked the *inter-rater agreement*, in order to see if there was a general consensus in the way the totality of the participants have rated the stereoscopic stimuli. We performed this test by computing *Kendall's coefficient of concordance* (explained in Appendix B.4) between every two sets of ratings corresponding to different participants. From these values, we were able to create a $N \times N$ matrix of correlation values for each of the 8 experiments, matrix in which the value 1 for the (i, j) position would mean strong correlation between the scores of participants i and j, and in which 0 would mean weak or no correlation between them. Such a matrix allowed us to easily identify the participants that correlated less with the rest and to discard them.

The number of participants discarded with each of the two approaches and also the number of participants kept after this phase are synthesized in Table 9.3. The table only shows however the numbers related to the experiments on the 3D TV screen.

Table 9.3: The number of discarded and valid participants for each sub-test of Experiment 2 and with each one of the two methods, *i.e.* the ITU method (ITU) and the method using Kendall's coefficient of concordance (K).

	tasks			
	image quality	comfort	realism	overall quality
discarded (ITU)	0	1	0	0
discarded (K)	5	2	6	2
valid	10	12	9	9

As for the tests on the 3D glasses, despite the fact that the ITU screening method did not discard any participant scores, the matrices of correlation clearly showed that the agreement between the majority of the pairs of any two participants was not sufficiently strong to indicate a general consensus. The results of this second screening test determined us to declare the quality ratings obtained in the 4 experiments on the prototype 3D glasses as not sufficiently accurate for exploitation. As a consequence, they were discarded.

Thus, only the results related to the test on the 3D TV screen will be discussed in the rest of this chapter. Interesting data on the 3D glasses will be nonetheless presented in this manuscript, but in another context, *i.e.* in the next chapter, where we discuss aspects related to the evolution of the physical symptoms of the observers while watching stereoscopic images.

9.1.4.5 Mean opinion scores and confidence intervals

As a last step in the experimental data processing chain, it was essential to concentrate all the results into a series of summary values. These would ease the interpretation of the experiment outcome and would allow drawing meaningful graphical representations of the condensed results.

To this end, we followed the ITU indications on the analysis and presentation of experimental results [ITU 2012b] and we computed the *mean opinion scores (MOS)* and the corresponding *confidence intervals* on various levels of the ratings matrices. Of most interest were the MOS per test condition levels, across all the participants and all the images. The formulas for computing these MOS and their corresponding confidence intervals are given in Appendix B.1 and the results obtained are discussed in the following section.

9.1.5 Results

In this section, the results of the first quantitative experiment are shown in the form of graphical representations of the MOS calculated in function of the test condition levels of each of the three independent variables that define the stimuli set and across all the other dimensions of the ratings matrices.

Thus, in the following figures, the leftmost graphical representation shows the MOS obtained for the three groups of images of different blur levels, the middle graphical representation illus-

trates the MOS for the three groups of images of different levels of the largest crossed disparities, and the rightmost one shows the MOS for the three sets of images taken from different capture distances.

9.1.5.1 Results on image quality

Figure 9.5 shows how the participants to our study rated the *image quality*, as it was described on the task sheet reproduced in Appendix D.2. These results show that differences in the judgment of the image quality were only triggered by the variations of *blur* that were present in the stimuli set and that the other two variables (crossed disparities and capture distance) did not generate any significant changes in the scores.

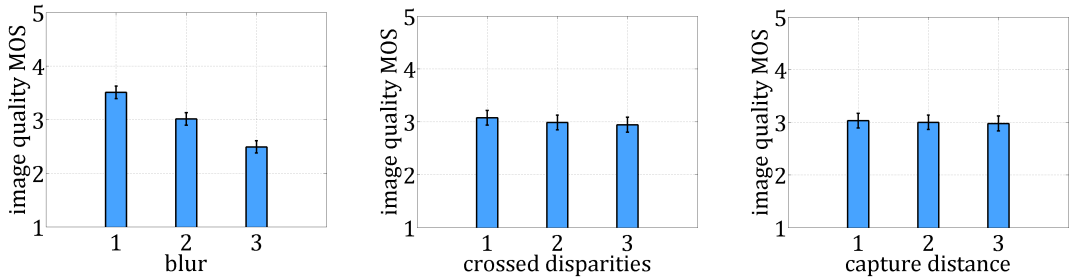


Figure 9.5: The image quality MOS in function of the three test conditions of each dataset parameter.

These observations, that can be visualized on the graphical representation, are supported by the results of the ANOVA tests that we performed in order to see if the means of the scores obtained for every two levels of the same test condition differ or not. All these statistical results are detailed in Table 9.4.

Table 9.4: The ANOVA test results for the scores obtained on *image quality* during Experiment 2; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	blur		crossed disparities		capture distances	
	F	p	F	p	F	p
1 – 2	**65.48**	$\ll$ **0.01**	0.55	0.4626	0.07	0.7857
2 – 3	**73.32**	$\ll$ **0.01**	0.16	0.6905	0.04	0.8402
1 – 3	**324.34**	$\ll$ **0.01**	1.28	0.2613	0.21	0.6484

9.1.5.2 Results on comfort

In the graphical representation of the MOS on *comfort* in Figure 9.6, both the *blur* variations and the *crossed disparities* seem to influence the comfort perceived.

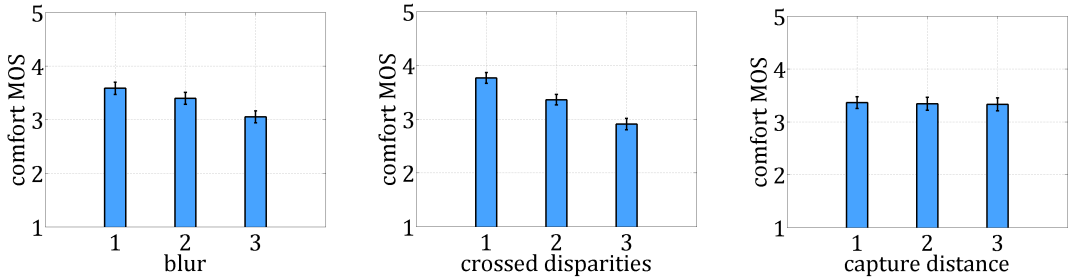


Figure 9.6: The comfort MOS in function of the three test conditions of each dataset parameter.

A closer look to the statistical results summed up in Table 9.5 shows, as expected, that large crossed disparities generate discomfort and that the three different levels of crossed disparities generate three statistical distinct groups of subjective reactions. This indicates that the comfort percept is gradual and tightly related to the disparities present in the scene.

Table 9.5: The ANOVA test results for the scores obtained on *comfort* during Experiment 2; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	blur		crossed disparities		capture distances	
	F	p	F	p	F	p
1 – 2	3.28	0.0746	**33.74**	$\ll$ **0.01**	0.04	0.8447
2 – 3	**12.56**	**0.007**	**34.23**	$\ll$ **0.01**	0.09	0.7687
1 – 3	**30.24**	$\ll$ **0.01**	**117.67**	$\ll$ **0.01**	0.01	0.9192

Moreover, statistical differences were found among the comfort ratings given to stereoscopic images of different levels of blur. However, not all the three categories of blurred images were judged as distinct from the point of view of the comfort they induce. Only the one corresponding to the strongest blur stood out from the two others, indicating that comfort is significantly deteriorated only when the level of blur is consistent.

9.1.5.3 Results on realism

When analyzing the results obtained after the *realism* task, we can notice that no influence of the capture distance variations was noticed in the scores. On the contrary, surprisingly, the *blur* and the *crossed disparities* variations seemed to influence the perceived realism, as it can be seen in Figure 9.7.

The statistical data obtained from the realism scores and reproduced in Table 9.6 confirm that the three MOS values corresponding to the stereoscopic images with different values of blur are statistically distinct. As for the crossed disparities variable, only the category of images with the largest negative disparities was determined as statistically different from the others, which shows that the influence of this factor on the perception of realism becomes noticed as negative only after a certain threshold.

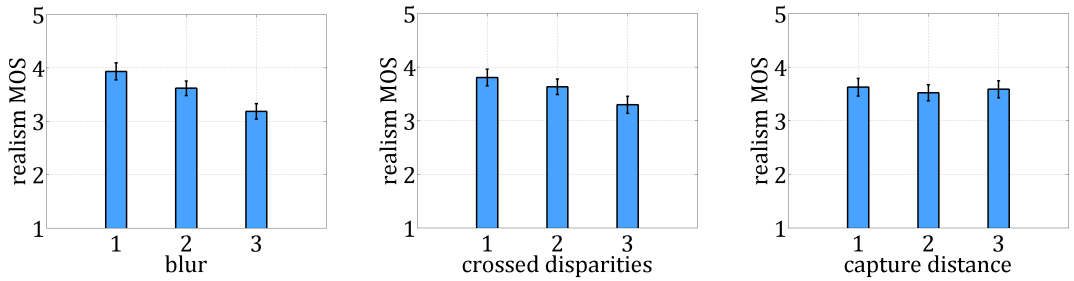


Figure 9.7: The realism MOS in function of the three test conditions of each dataset parameter.

Table 9.6: The ANOVA test results for the scores obtained on *realism* during Experiment 2; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	blur		crossed disparities		capture distances	
	F	p	F	p	F	p
1 – 2	**14.11**	**0.0004**	3.06	0.0845	0.88	0.3523
2 – 3	**34.18**	**$\ll$ 0.01**	**13.04**	**0.0006**	0.37	0.5467
1 – 3	**84.11**	**$\ll$ 0.01**	**26.72**	**$\ll$ 0.01**	0.12	0.733

9.1.5.4 Results on the overall stereoscopic quality

In the last set of data, shown in Figure 9.8 a strong influence of *blur* on the *overall stereoscopic quality* ratings can be noticed. The different *crossed disparities* also generated different perceptions of the overall impression of quality, while the capture distance still did not have any effect on the ratings of the participants.

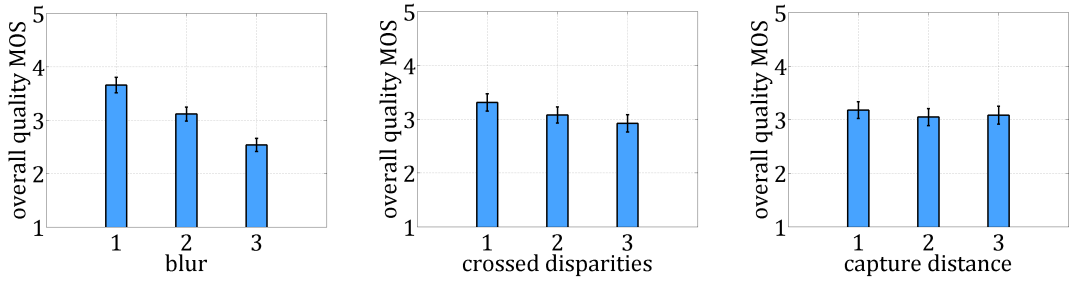


Figure 9.8: The overall stereoscopic quality MOS in function of the three test conditions of each dataset parameter.

These conclusions were confirmed by the statistical tests, detailed as well in Table 9.7, which showed that the three categories of images with different variations of blur are all statistically different and that only the two extreme categories of images with different crossed disparities generated statistically different MOS.

Table 9.7: The ANOVA test results for the scores obtained on the *overall stereoscopic quality* during Experiment 2; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	blur		crossed disparities		capture distances	
	F	p	F	p	F	p
1 – 2	**46.34**	$\ll$ **0.01**	2.99	0.0881	1.04	0.3124
2 – 3	**60.81**	$\ll$ **0.01**	1.69	0.1972	0.07	0.7984
1 – 3	**223.47**	$\ll$ **0.01**	**9.31**	**0.0032**	0.56	0.4558

9.1.6 Conclusions after Experiment 2

As shown in Section 9.1.1, the dataset that we used for this first experiment had a structure that was based on several hypotheses of correlation between the physical stereoscopic image parameters and the perceptual attributes previously identified during our qualitative study. Therefore, the first verification to be done on the results of this experiment was to check whether our hypotheses were correct.

Two of the expected correlations were successfully validated, *i.e. the influence of the blur on the image quality perceived* and *the influence of the crossed disparities on comfort.*

However, *no variation in the realism ratings* has been observed *in function of the variation of the capture distances* considered. We found several explanations for our results.

First, we concluded that the naïve observers might be unfamiliar with artifacts particular to stereoscopic content, like the cardboard effect, that they have difficulties in identifying, especially if its appearance is not very pronounced and if the other alterations brought to the stereoscopic images are more familiar to them and easily identifiable.

Second, we think that the difficulty in distinguishing the geometrical distortions was also due to the short duration of a stimulus display, in contrast with the previous exploratory study, where the participants identified the cardboard effect, but after watching the images for as long as they wanted.

Third and last, we considered that the choice of natural scenes for our dataset influenced our results. The same is valid for the fact that we tried to outline the cardboard effect using a stereoscopic capture system with a baseline close to the human interocular distance and a small focal distance. Previous studies emphasized the cardboard effect using large focal depths, while photographing a strictly controlled scene, with a single object and no other depth cues [Yamanoue 2000]. This case is opposed to our data set, where a large number of depth cues were present in each of the chosen scenes. We bring into discussion the depth cues, since we consider that the complex human perception can use them for compensating a slightly altered rendering of depth.

Other more complex, unexpected, dependencies between the physical stereoscopic image parameters and the perceptual attributes assessed were identified by this experiment as well. The results have shown that *the comfort perception can also be correlated to image quality artifacts,*

like blur, but also that, in the absence of other parameters impacting on it, *the impression of realism was influenced both by blur variations and by crossed disparities variations.*

Also, since no influence of the capture distance variations was observed for any of the image quality, comfort, or realism perceptual attributes, we were not surprised to see that this physical parameter did not influence on the overall stereoscopic quality percept either.

These conclusions managed therefore to bring answers to the questions asked at the beginning of the experiment and illustrated in Figure 9.1. These answers, schematized in Figure 9.9, are, on one hand, the confirmation of two of the hypothesized correlations between the physical parameters and the perceptual attributes (in red) and, on the other hand, the demonstration of some other unexpected such correlations (in blue). The third hypothesis related to the realism percept was not confirmed by this experiment (represented in black), as explained.

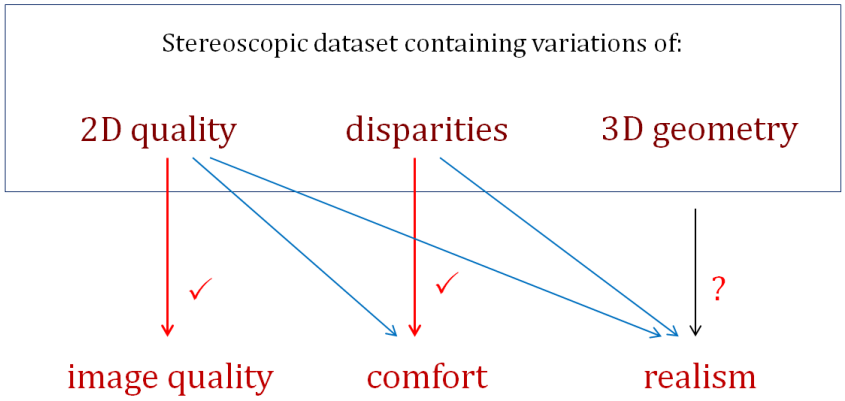


Figure 9.9: The adaptation of our initial stereoscopic image quality model after Experiment 2.

Moreover, the large set of subjective data collected during this experiment constitutes, in Chapter 11, the *numeric support* for the elaboration of *the mathematical structure of our stereoscopic image quality model.*

9.2 Experiment 3

The fact that the hypothesis on the correlation between the capture distance parameter and the realism perceived was not confirmed during our first quantitative experiment led to a missing link in the validation of our proposed stereoscopic image quality model described by Figure 8.20. Therefore, a second quantitative experiment was designed in order to shed more light on the realism percept, by using a new stereoscopic dataset with other physical parameters variations. The problem posed by this second experiment is sketched in Figure 9.10.

9.2.1 The choice of the stimuli

The new stimuli set designed for Experiment 3 was the *MOOV3D – GIPSA Stereoscopic Dataset 3.* It was created with an exclusive focus on the parameters that influence the cardboard effect

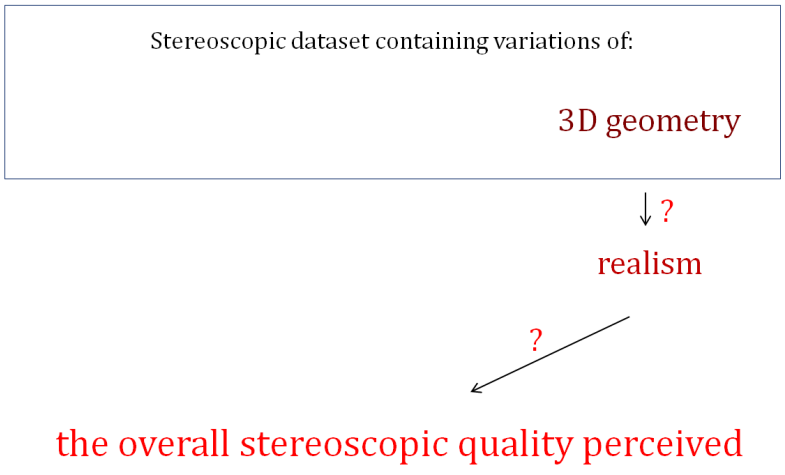


Figure 9.10: The objectives of Experiment 3.

and which could therefore induce variations in the perception of realism during our second quantitative experiment. Among all these parameters, we considered this time the focal distance and the position of the photographed object in space, relative to its background, which justifies the dataset structure synthesized by Table A.5.

Our first formulated hypothesis was that larger *focal distances* would induce a stronger *cardboard effect*. We based this hypothesis on the analysis of the geometry of our stereoscopic system, which showed that the cardboard effect took considerably larger values when using the maximum focal distance allowed by our capture device ($18.9\,mm$), than when using its minimum focal distance ($6.3\,mm$). The graphical representation of this observation is shown in Figure 9.11 and the calculations were performed according to Appendix E.3.

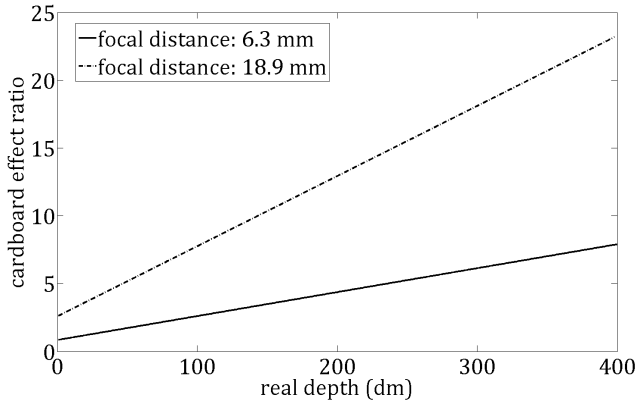


Figure 9.11: The evolution of the cardboard effect with the capture distance for two focal distances.

Our second hypothesis was that larger *distances between the photographed object and its background* would emphasize this effect. We decided to include variations of this parameter as a consequence of our exploratory study, where among the stereoscopic images that induced the cardboard effect were images with such large object-background distances.

In this form, Dataset 3 was the stimuli set used in our second quantitative experiment, allowing us to verify if, this time, the considered physical parameters influence the realism percept.

9.2.2 Experimental protocol and participants

Experiment 3 also used the *quantitative protocol* described in Section 7.3 and was performed exclusively on the 3DTV described in Section 6.3.

The participants to this experiment only had two concepts to rate and these were *realism* and *overall stereoscopic quality*, both again described as in Appendix D.2. Therefore, Experiment 3 was only made up of 2 sub-experiments and its simple structure is schematized in Figure 9.12.

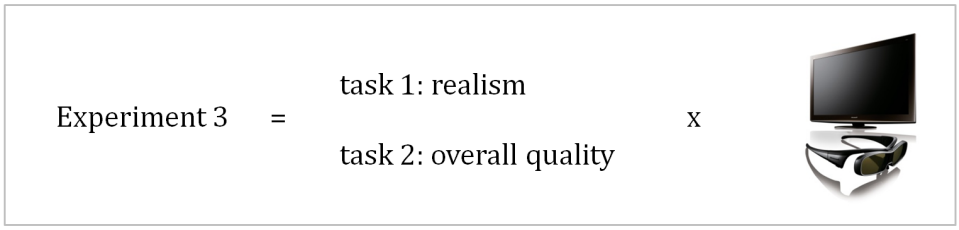


Figure 9.12: The structure of Experiment 3.

The number of valid participants that performed the tasks was 20 for the test on the realism and 21 for the test on the overall stereoscopic quality.

The detailed test plan of Experiment 3 is reproduced in Appendix D.2.

9.2.3 Data collection and processing

The data collection and data processing phases have been performed in the same manner as for Experiment 2, with just a few differences of parameters. At the end of Experiment 3, 41 text files had been collected, while the 2 matrices of ratings obtained were of size $N \times TC1 \times TC2 \times K \times R$, with the same notation conventions, except for:

$TC1$: the number of test conditions for the first variable (focal distance);

$TC2$: the number of test conditions for the second variable (object-background distance).

The actual values of the parameters of Experiment 3 for each of its 2 sub-experiments were:

N: 20 (realism) or 21 (overall stereoscopic quality);

$TC1$: 3;

$TC2$: 3;

K: 8;

R: 2.

This consequently led to two $N \times 3 \times 3 \times 8 \times 2$ results matrices.

The observations made in Section 9.1.4.2 on the scores distribution for Experiment 2 are equally valid for this experiment as well, in the sense that the shapes of the histograms of scores for the 2 tests of Experiment 3 are both *platykurtic*, and both sets of ratings collected can be considered *normally distributed* in the statistical studies to be performed on the data. The 2 histograms of votes and their corresponding kurtosis value are displayed in Figure 9.13.

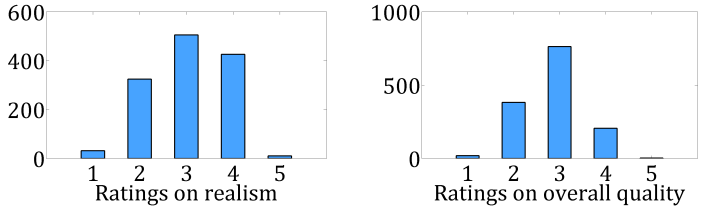


Figure 9.13: The histograms showing the scores distribution for each of the 2 tests of Experiment 3.

Table 9.8: The kurtosis measures of the score distribution per task and per display during Experiment 3.

	realism	overall quality
3D TV screen	2.1883	2.4114

The normalization of the experiment scores has been equally done by computing the corresponding standard scores and by remapping these scores to the interval of our original scale.

Regarding the participant screening, a summary is presented in Table 9.9, showing the number of participants that were discarded for each of the 2 tests of Experiment 3.

Table 9.9: The number of discarded and valid participants for each sub-test of Experiment 3 and with each one of the two methods, *i.e.* the ITU method (ITU) and the method using Kendall's coefficient of concordance (K).

	tasks	
	realism	overall quality
discarded (ITU)	1	2
discarded (K)	5	6
valid	12	11

And the MOS and their corresponding confidence intervals have been computed as well as explained for Experiment 2, in Section 9.1.4.5.

9.2.4 Results

Similarly to Experiment 2, the results of Experiment 3 are detailed in the form of the MOS computed in function of the test condition levels of the two independent variables defining Dataset 2 and across all the other dimensions of the ratings matrices. In the following figures, the left representation shows the MOS for the three groups of images taken with different focal distances, while the right one gives the MOS for the images grouped in function of the distance between the main object and the background at capture.

9.2.4.1 Results on realism

Figure 9.14 shows how the MOS obtained after the realism task differ considerably among the three categories of images taken with distinct *focal distances*. No variation in the rating results was observed however in the scores grouped in function of the different object-background distances of the stereoscopic images.

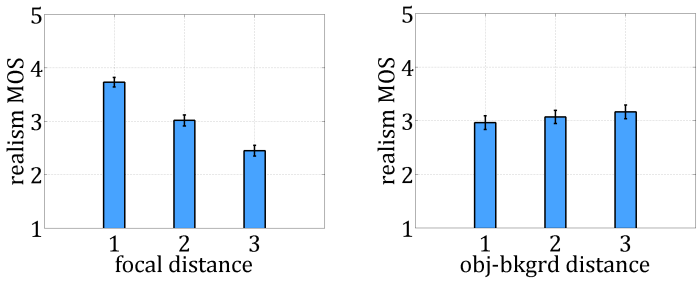


Figure 9.14: The realism MOS in function of the two test conditions of each dataset parameter.

The detailed results of the statistical tests for this experiment can be found in Table 9.10 and they confirm the strong influence of the focal distance on the realism perceived by the participants.

Table 9.10: The ANOVA test results for the scores obtained on *realism* during Experiment 3; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	focal distance		object-background distances	
	F	p	F	p
1 – 2	**93.49**	$\ll$ **0.01**	0.41	0.523
2 – 3	**59.25**	$\ll$ **0.01**	0.32	0.577
1 – 3	**337.21**	$\ll$ **0.01**	1.38	0.246

9.2.4.2 Results on the overall stereoscopic quality

The results on the overall stereoscopic quality follow approximately the same trends as in the case of the results on the realism, as shown in Figure 9.15, in the sense that a strong influence of the focal distance variations on the realism perceived can be observed, whereas no subjective variations were generated with changes on the object-background distances.

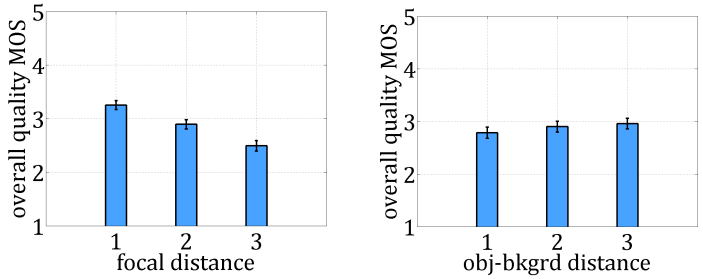


Figure 9.15: The overall stereoscopic quality MOS in function of the two test conditions of each dataset parameter.

The statistical differences among the groups of stereoscopic images photographed with different *focal distances* were confirmed by the statistical measures summarized in Table 9.11.

Table 9.11: The ANOVA test results for the scores obtained on the *overall stereoscopic quality* during Experiment 3; the TC column stands for the test conditions compared for each one of the three perceptual attributes; the values in bold are the statistical significant results.

TC	focal distance		object-background distances	
	F	p	F	p
1 – 2	**25.06**	$\ll$ **0.01**	0.84	0.3649
2 – 3	**19.16**	$\ll$ **0.01**	0.21	0.6489
1 – 3	**69.15**	$\ll$ **0.01**	1.88	0.1772

9.2.5 Conclusions after Experiment 3

The second quantitative experiment that we implemented was a continuation of Experiment 2. It had the purpose of bringing clarifications to one of the three hypothesis that Experiment 2 did not manage to validate, and this was the hypothesis referring to the perception of the *realism* attribute.

Experiment 3 used a different dataset in order to verify if other two factors than those tested in Experiment 2 could influence the realism perceived. Two initial hypotheses were thus formulated during this second experiment on the influence of these two factors on the judgments on realism and Experiment 3 tested them.

Therefore, the main result that stood out from the processing of the data of Experiment 3 was related to our initial hypotheses of this second experiment. One of the two variables that

we initially considered, *i.e.* the *focal distance*, proved to influence the *realism* perceived and, moreover, also had an influence on the *overall stereoscopic quality*.

The influence of the second parameter – *the position of the main object in the photographed scene* relative to its background – on the percept of realism *has not been validated*. A possible reason for this could be the choice of the object-background distances, which were probably too close to generate significant variations in the judgment of the realism.

These findings are illustrated by Figure 9.16, which shows that the questions asked by Experiment 3 were unambiguously answered in its results.

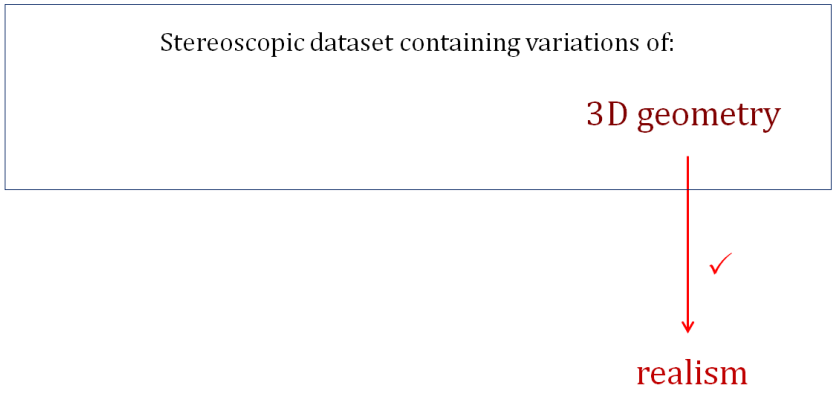


Figure 9.16: The conclusions of Experiment 3.

Like for Experiment 2, the totality of the subjective data collected during Experiment 3 is of great value, since it contributes as well to *the mathematical refinement of our stereoscopic image quality model*, as Chapter 11 explains.

Conclusions

In our investigation on how human observers perceive and judge the stereoscopic image quality, *two quantitative experiments* have been implemented, following the results of a previous qualitative exploration. Their purpose was to *validate* and *bring more precision* to the first *stereoscopic image quality model* that we sketched after the qualitative study. This model to validate is illustrated in Figure 8.20.

To this end, we created two stereoscopic datasets with precise structures. These datasets were assessed by human observers in order to gather their judgments on the *overall percept of stereoscopic image quality*, but also on three perceptual attributes that we found to define the global percept: *image quality*, *comfort*, and *realism*. The questions raised by our quantitative studies were thus formulated as shown in Figure 9.1.

The results of the two investigations first confirmed the relations between a series of physical stereoscopic image quality parameters and the evaluated perceptual attributes. The scores on the *image quality* attribute were determined to be related to the presence of a *2D quality impairment* – *blur*. The perception of *comfort* varied with the levels of the largest *crossed disparities*, but

also with the levels of *blur*. And the influence of the *3D geometry distortions*, determined by different *focal distances*, on realism was also proven, with *blur* and *crossed disparities* influencing this perceptual attribute as well. Figure 9.17 concentrates these findings, showing in red all our hypothesis that were confirmed and in blue other unexpected relations that we could identify.

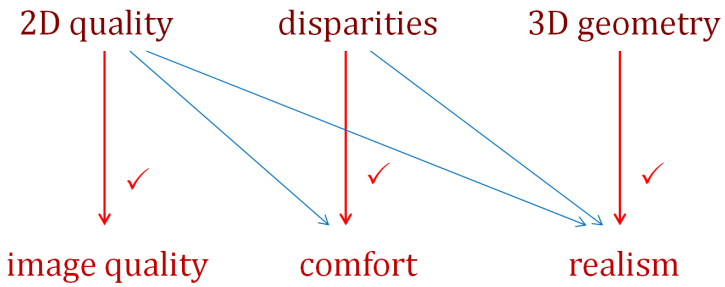


Figure 9.17: The conclusions of our quantitative study.

As a consequence, these results have proven that the *datasets* that we elaborated could successfully generate *variations in the three perceptual attributes mentioned* and, moreover, that these variations were *also reflected in the overall stereoscopic quality perceived.*

Given this important confirmation, the central problem of our work was ready to be addressed: finding the right mathematical framework which groups together all the relations observed so far.

This last step in our work will be detailed by Chapter 11, where the numerous sets of data collected during Experiments 2 and 3 have constituted an important numerical support for the mathematical refinement of our stereoscopic image quality model proposed.

Chapter 10

A Human Factors Study

Introduction

The previous chapters have shown how the human judgment on the stereoscopic image quality could be assessed through precisely designed experiments, then structured in evaluation frameworks. The exact mathematical modeling of the perceived stereoscopic quality will be presented in the final chapter of this manuscript, Chapter 11.

Before that, however, in order for our view on the impact that the stereoscopic images have on the observers to be broader, we considered of essential value to also assess the state of well being of the participants during our subjective tests.

To this intent, along with the first quantitative experiment presented in the previous chapter, we also performed an investigation of the evolution of various physical symptoms of the participants during the visualization of stereoscopic images. This investigation is what constitutes *our human factors study* and will be comprehensively described in the present chapter.

10.1 Context

The human factors study presented in the following sections was implemented in the context of our first quantitative experiment with the purpose of understanding the influence of factors like the type of the display, the level of artifacts, or the visualization task on the human symptoms.

Thus, the larger frame of this study was that of the *quantitative evaluation* in Section 9.1, which was composed of 8 sub-experiments. Human observers watched stereoscopic still images

containing different *variations of the content parameters*, displayed on two different devices – the ultra-wide *3D TV screen* or the pair of prototype immersive *3D glasses* described in Section 6.3 –, and had to rate them in function of their preference for one of four concepts previously described: *image quality*, *comfort*, *realism*, and *overall stereoscopic quality*.

10.2 Experiment protocol and participants

For implementing this study, we followed closely the *human factors protocol* described in Section 7.4, which is centered on the *Simulator Sickness Questionnaire (SSQ)*.

Since the SSQ was applied in the form of a computer questionnaire before and after each of the 8 tests of Experiment 2, its participants were those described in Section 9.1.2, distributed per task and per display as shown in Table 9.1.

10.3 Data processing

The raw results obtained after this study were 8 collections of scores from 0 to 4, one collection for each of the 8 tests in Experiment 1. Every such collection contained two structures of $N \times 12$ scores, corresponding to the symptom scores given by the N participants of the corresponding test before and after the visualization experiment.

Before the processing of these data, we performed the screening of the participants, but differently than for our quantitative studies. We discarded the data corresponding to any participant that had simulator sickness symptoms that were too strong before the test, *i.e.* scores larger than 1 for any of the questions. The final number of valid participants after this screening is given in Table 10.1, with the original number of participants accepted after the vision tests in brackets. We did not considered a screening related to the general consensus necessary, since we are aware that the physical symptoms can vary a lot among participants and since what interested us was the general trend, computed from the totality of our data.

Table 10.1: The number of valid participants for each sub-experiment; in brackets, the number of participants discarded.

	tasks			
	image quality	comfort	realism	overall quality
3DTV	15 (0)	11 (4)	12 (3)	10 (1)
3D glasses	12 (1)	10 (2)	11 (0)	10 (0)

The data processing of the remaining results was performed according to the method presented in the original article proposing the SSQ [Kennedy 1993], method that is reproduced in Appendix C.

The method consists in weighting, then summing up the scores into three main values, representing three clusters of symptoms: N for the *nausea* cluster, O for the *oculomotor* cluster,

and D for the *disorientation* cluster. A *total score* was also computed, noted as TS. The weights used for these calculations are given in Table 7.1.

10.4 Results and observations

The concentrated results of our SSQ study are illustrated in Figure 10.1, where the differences between the pre-exposure and post-exposure symptoms for both displays and for each of the four visualization tasks can be observed. The figure also shows the statistical level of significance of these differences, computed using the Wilcoxon test, which is explained in Appendix B.7. All these statistical results are synthesized in Table 10.2.

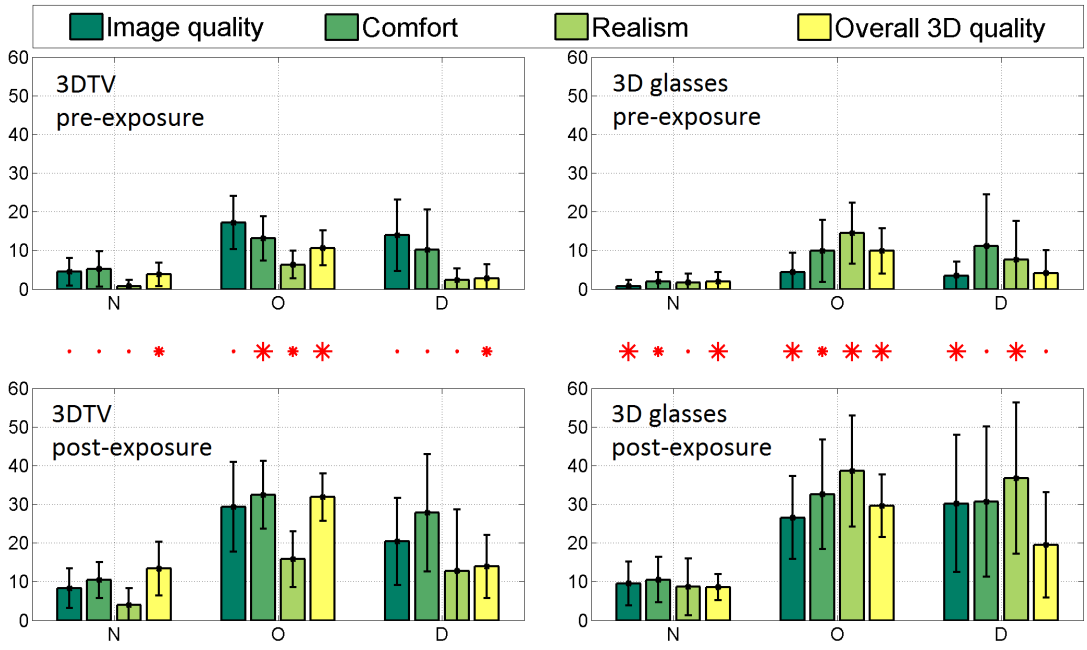


Figure 10.1: The N, O, and D cluster symptom scores for the 3D TV screen (left) and the 3D glasses (right), as recorded before (top) and after (bottom) the visualization experiments for each of the four tasks (in different colors). The * symbols indicate the significance levels of the Wilcoxon rank sum test on the differences between corresponding categories before and after the test (1 %, 5 %, and >5 % significance level from the largest to the smallest * symbol).

10.4.1 Results on the 3D TV screen

After the visualization experiments on the 3D TV screen, no symptom differences were observed for the *image quality task*. This can be easily justified by the high resolution and large size of the screen, on which subtle differences of blur can be distinguished with no considerable effort.

For the *comfort task*, however, an important augmentation of the oculomotor symptoms was recorded ($p = 0.0034$). This can be explained by the fact that the database contained images with crossed disparities close to the comfort limit of the human visual system. Therefore, when the participants focused on rating this task, they were particularly attentive and sensitive to the

fact that the visual exploration was demanding in terms of convergence and accommodation.

Table 10.2: The Wilcoxon test results computed on the SSQ scores obtained during Experiment 2; each number corresponds to the p-value produced by the comparison of the pre-exposure and post-exposure symptom scores for each task and for each display; the values in bold are the statistical significant results.

display	N				O			
	image quality	comfort	realism	overall quality	image quality	comfort	realism	overall quality
3D TV	0.3216	0.1295	0.2660	**0.0483**	0.1198	**0.0034**	**0.0411**	$\ll$ **0.01**
3D glasses	**0.0096**	**0.0207**	0.1296	**0.0093**	$\ll$ **0.01**	**0.0210**	**0.0074**	**0.0019**
display	D				TS			
	image quality	comfort	realism	overall quality	image quality	comfort	realism	overall quality
3D TV	0.4709	0.0764	0.3037	**0.0244**	0.2614	**0.0382**	**0.0363**	**0.0011**
3D glasses	**0.0024**	0.0521	**0.0094**	0.0579	$\ll$ **0.01**	**0.0316**	**0.0062**	**0.0026**

The *realism task* led as well to higher post-exposure oculomotor scores ($p = 0.0411$), but the largest human factor changes for the 3D TV screen experiment were recorded for the *overall stereoscopic quality task*: statistical differences have been determined between the pre- and post-exposure scores for the nausea ($p = 0.0483$), oculomotor ($p << 0.01$), and disorientation ($p = 0.0244$) symptom clusters. This happened because rating the overall stereoscopic quality task meant taking into account at once several factors related to image quality, comfort, and realism. Therefore the visualization effort involved by this multi-dimensional task can be seen as an implicit cumulative effort of all the other three tasks.

10.4.2 Results on the 3D glasses

In the case of the tests on the 3D glasses, a larger number of changes than for the 3DTV were observed between pre- and post-immersion.

The *image quality task* generated an increase in all the three symptom clusters: nausea ($p = 0.0096$), oculomotor ($p << 0.01$), and disorientation ($p = 0.0024$). In contrast to the results on the 3D TV screen, this shows that on the micro size screens of smaller resolution of the prototype 3D glasses, distinguishing subtleties of image quality can be more demanding.

The *comfort task* on the 3D glasses augmented the oculomotor symptoms ($p = 0.0210$), but less than when it was performed on the 3D TV screen. The result is in accordance with the geometry of the 3D glasses, for which the representation of crossed disparities was less aggressive for the human visual system than the 3D TV screen configuration considered for our test. The comfort task on the prototype 3D glasses has increased however the nausea symptoms as well, which gives a hint of the fact that an exploration with focus on the different levels of disparities, given the immersion of the 3D glasses, can have consequences on the general state of well being.

After the *realism task*, a considerable deterioration was observed in the scores on the oculomotor ($p = 0.0074$) and disorientation ($p = 0.0094$) symptoms. For explaining this result, we mention that the ratings the participants gave for the realism attribute during the test on the prototype 3D glasses were lower than those during the test on the 3D TV screen. Therefore, we can conclude that an exploration in order to decide whether a 3D image is similar to what can be seen in reality is more strenuous when this resemblance can be found more difficultly.

The *overall stereoscopic quality task* lead to changes in the nausea ($p = 0.0093$) and oculomotor ($p = 0.0019$) symptom clusters, following the same logic as for the 3D TV screen. No important changes of disorientation were observed after this task ($p = 0.0579$) on the 3D glasses.

10.4.3 Total scores

The total scores that concentrate all the results presented so far are given in Figure 10.2. This graphical representation sums up the fact that stronger symptoms have been recorded after the tests performed on the prototype 3D glasses, and also that all the tasks have induced a deterioration of the symptoms using the glasses, compared to only three of the tasks when watching the 3D TV screen.

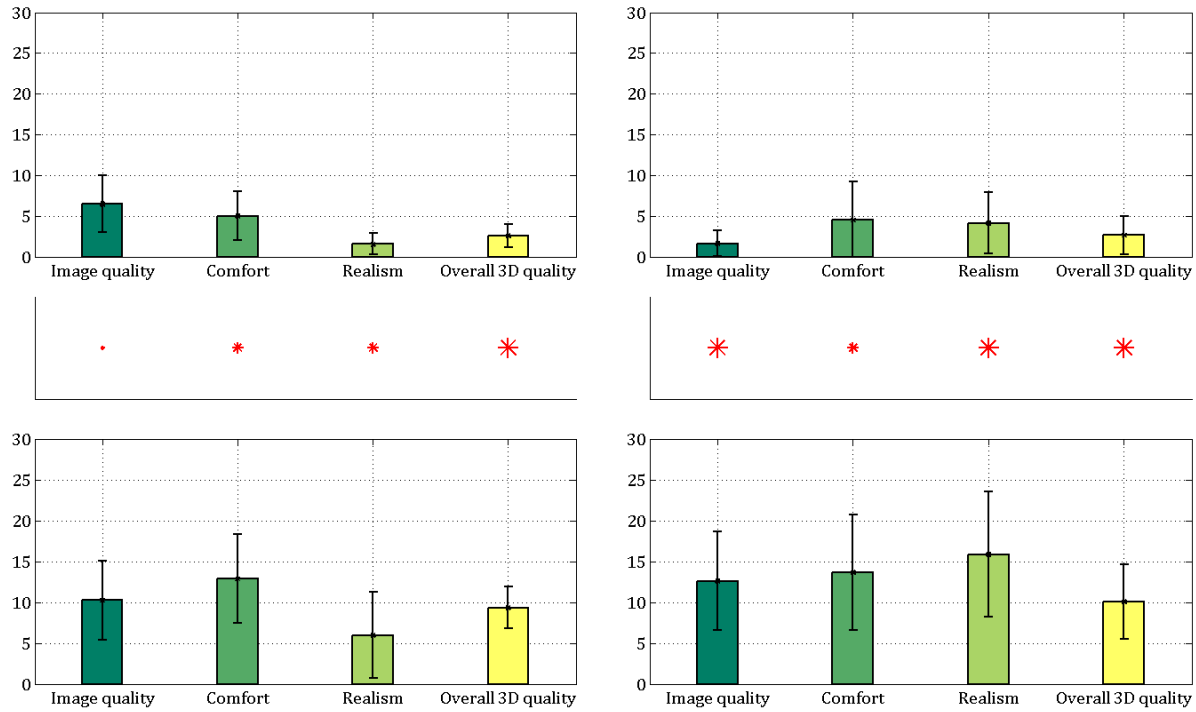


Figure 10.2: The total scores for the 3D TV screen (left) and the 3D glasses (right), as recorded before (top) and after (bottom) the visualization experiments for each of the four tasks (in different colors). The * symbols indicate the significance levels of the Wilcoxon rank sum test on the differences between corresponding categories before and after the test (1 %, 5 %, and >5 % significance level from the largest to the smallest * symbol).

10.4.4 Detailed results

In order to better understand what led to this results, we continued our analysis at each symptom level.

We could observe that there were a series of questions for which only the experiments on the prototype 3D glasses brought changes in the post-exposure symptoms: the image quality task generated more general discomfort, more fatigue, and more difficulty focusing; the comfort task led to augmented scores for the headache; the realism deteriorated the general comfort and the capacity to focus, and augmented the blurred vision and the dizziness with the eyes open.

However, on the 3D TV screen the comfort task increased the fatigue and the overall 3D quality task augmented the eyestrain.

10.4.5 Interpretation of the results

The augmentation of certain symptoms in the case of an HMD used as a personal screen is in line with the findings in [Howarth 1997] and this could be an indicator of the fact that the HMD devices still leave place for improvements in visualization comfort. In the mean time, it is expected that watching stereoscopic content in a completely immersive environment like an HMD, but also *e.g.* in a movie theater, would cause more simulator sickness symptoms compared to watching the same content on a TV screen, where the natural ambient vision is still present. For this reason, for typical commercial stereoscopic content like 3D movies, there is usually a post-processing phase where the stereoscopic depth is adapted to the scene and reduced most of the time, while quality artifacts are avoided, in order to prevent phenomena related to visual fatigue.

In our particular test case, we did not use such improved content, but specific still images containing various artifacts, with the purpose of understanding the impact of stereoscopic images on the public when the optimal quality of the content fails to be reached.

Also, a physiological adaptation to this type of immersive visualization was found in [Jumisko-Pyykko 2010b], while an adaptation between successive sessions of testing was noted in [Kennedy 1993], therefore the lack of familiarity of the consumers with a personal display type of system might be another explanation for our results.

Conclusions

In the study presented in this chapter, we considered a *consumer situation* for which we tested the acceptance, from the human factors perspective, of two different stereoscopic displays, from which one was a prototype, for tasks of stereoscopic image visualization. We were mostly interested in better understanding what aspects influence the symptom evolution during the stereoscopic visualization.

The results allowed us to conclude that *the task* performed during the visualization has a considerable influence on the evolution of the simulator sickness symptoms. Also, we found that *different stereoscopic displays* generated *different symptoms*, in function of *the content displayed*.

This set of results allowed us to consolidate our knowledge about the perception and judgment of the stereoscopic image quality, by acquiring detailed conclusions on the human factors dimension.

Chapter 11

Towards A Method of Assessing the Stereoscopic Image Quality

Introduction

The previous chapters have shown a series of experimental results acquired on the relations between several physical stereoscopic image parameters (*2D quality parameters*, *disparities*, and *3D geometry parameters*) and the perceptual attributes identified as important in the assessment of the stereoscopic image quality (*image quality*, *comfort*, and *realism*).

This last chapter concentrates on the mathematical frame that could group our previous findings. It has multiple purposes: to propose a *perceptual stereoscopic image quality model*; to illustrate how *visual algorithms* can *explain the perceptual attributes* mentioned, given the *physical image parameters*; and to understand how the *overall percept of stereoscopic image quality* can be directly estimated from these *physical image parameters*.

The following sections first make a summary of the data under study, then present the mathematical tools used in our approach, the results that we obtained, and the interpretation that we gave to these results.

11.1 The data

Table 11.1 sums up the data accumulated through our subjective experiments and through our algorithmic processing of the stereoscopic datasets created.

Table 11.1: Our collection of subjective and objective data on the stereoscopic image quality percepts and parameters.

	subjective scores				objective scores		
Experiment 2	IQ_2	C_2	R_2	SIQ_2	*blur*	*cr_disp*	×108
Experiment 3			R_3	SIQ_3	*cdb_eff*		×72

On one hand, 4 sets of 108 subjective scores are available, one for each of the following perceptual attributes: *image quality* (IQ_2), *comfort* (C_2), *realism* (R_2), and *overall stereoscopic quality*, also called *stereoscopic image quality*, (SIQ_2). These have been obtained during Experiment 2 and correspond to the stereoscopic images in the MOOV3D – GIPSA Stereoscopic Dataset 2, as described in 9.1. Also, 2 other sets of 72 subjective scores exist. They have been attributed to the perceptual attributes *realism* (R_3) and *overall stereoscopic quality* (SIQ_3) during Experiment 2, on the MOOV2D – GIPSA Stereoscopic Dataset 3.

All these 6 sets of *mean opinion scores* constituted our *subjective data collection*.

On the other hand, a series of *objective measures* were computed as well.

First, the *blur level* that was present in our images were determined using the blur metric described in Appendix E.1. The blur measures for the dataset used in Experiment 2 thus led to scores distributed between 0.2932 and 0.6414 (where 0 stands for sharp images and 1 stands for images with strong blur), thus showing the variability of the blur level in this dataset.

Second, like for our exploratory experiment, the *largest crossed disparities* were computed for all our stereoscopic images, based on the results of a stereo matching algorithm using the SIFT feature detection method. This approach is described in Appendix E.2. The largest crossed disparities computed for the dataset used in Experiment 2 varied from -4.7984 cm to -0.0578 cm for the viewing distance of 1.5 m. This is equivalent to crossed disparities of up to 109.8 minutes of arc in absolute value, which, as justified in Section 2.1.5, were expected to overtake the threshold of comfortable viewing.

The *cardboard effect*, however, was not computed by directly analyzing the images, but was estimated at a theoretical level. Since the cardboard effect is a geometric distortion that can only be observed in relation to the original real-world geometry of the photographed scene, computing it without any information on the capture conditions would be a very complex endeavor, if not impossible. In our work, we were able to estimate a measure of the cardboard effect in our stereoscopic images by using the formulas described in Appendix E.3, because we knew all the parameters of the stereoscopic system that we used at capture. Like this, the cardboard effect ratio of a stereoscopic image was determined relative to the main object in the scene and was computed in function of the distance from the camera to this main object at capture, for every given focal distance.

Concretely, for the dataset used in Experiment 3, for each focal distance, we used approximately the same capture distance to capture all the scenes, which led, according to Appendix E.3, to theoretically estimated cardboard effect ratios from 1.3713 (corresponding to the 6.3 mm focal distance – almost no distortion) to 7.6721 (corresponding to the 18.9 mm focal distance –

a large compression of the depth planes). Since the calibration performed for the intermediate focal distance of 13.1 mm was less precise, we discarded the values in between these thresholds and only continued our study with the values corresponding to the limits of our interval of variation of the cardboard effect.

This method of estimating the cardboard effect from the theoretical perspective could also be replaced by a method that measures the actual cardboard effect ratio from a stereoscopic pair. Such a method would also need information on the real-world proportions at capture and would compare these proportions with the computed rendered proportions, to check whether they are maintained or of how much they were altered. The principles of applying such a measure are illustrated in Appendix E.3. Unfortunately, although we were able to test this method later on some sample images outside the frame of our experiments, at the moment of the capture of our datasets we had not stored all the necessary real-world proportions in order to be able to apply the method on the data of Experiment 2 or 3.

The three *objective measures* we enumerated are denoted by *blur*, *crossed_disp*, and cdb_eff in the formulas that follow and the objective scores they generated constituted our *objective data collection.*

We mention here that, since no variations in the perception of the cardboard effect have been detected in the structure of the dataset used in Experiment 2, we consider the *cardboard effect* of these 108 stereoscopic images as a *constant.* Similarly, at the creation of the dataset for Experiment 3, special attention was payed to having uniform *blur* across all the images, thus the blur can be considered a *constant* for Experiment 3. The disparities of the stereoscopic images in Dataset 3 were also computed. They were either all positive for the majority of the images, or the few crossed disparities determined were very small (up to -0.3754 cm for the viewing distance of 1.5 m; equivalent to a maximum of 22.5 minutes of arc in absolute value). Since only crossed disparities larger than approximately 30 minutes of arc proved a negative influence on the perceptual attributes assessed in our experiments, we considered that the physical parameter *disparities* could be considered a *constant* as well in the context of Experiment 3.

As a consequence, the summary of the objective scores and of their distribution in function of the experiment during which they were obtained is in Table 11.2.

Table 11.2: A summary of the objective measures.

	Experiment 2		Experiment 3	
	min	max	min	max
blur measure	0.2932	0.6414	constant	
crossed disparities (cm)	0.0578	4.7984	constant	
cardboard effect ratio	constant		1.37	7.67

All the data presented so far, subjective and objective, constitute the numeric basis on which all the studies presented in this chapter are based.

11.2 Our stereoscopic image quality model

Our approach on determining the mathematical form of the perceptual stereoscopic image quality model that we proposed in Figure 8.20 is actually based on finding the answer to the questions in Figure 11.1. We say *perceptual* because, as formulated in the SIQC[1], in Section 5.2.1, a stereoscopic image quality model is a formalism that connects the various *perceptual attributes* to the *overall percept* on the stereoscopic image quality. Therefore, this part of our mathematical study only focuses on modeling the connections among the subjective percepts evaluated by our participants.

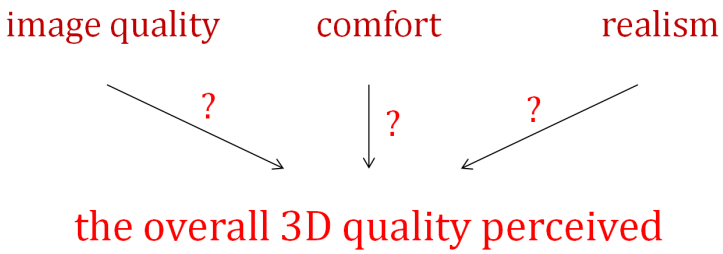


Figure 11.1: The central problem of our work formulated after the quantitative experiments and answered in Chapter 11.

Since, in concrete terms, our exact purpose was to explain the *stereoscopic image quality (SIQ)* in function of the *image quality (IQ)*, the *comfort (C)*, and the *realism (R)* perceptual attributes, for this study we used the data from Experiment 2, which contained all the needed subjective ratings.

In order to determine the best mathematical model that could fit our data, we first performed individual studies on the relations between the scores on the overall percept SIQ_2 and the scores obtained for each one of the component perceptual attributes measured, IQ_2, C_2, and R_2. Thus, by plotting SIQ_2 in function of the scores on these attributes, we could conclude that there was a *linear influence* of each perceptual attribute on the overall scores. The scatter plots that show these relations are in Figure 11.2.

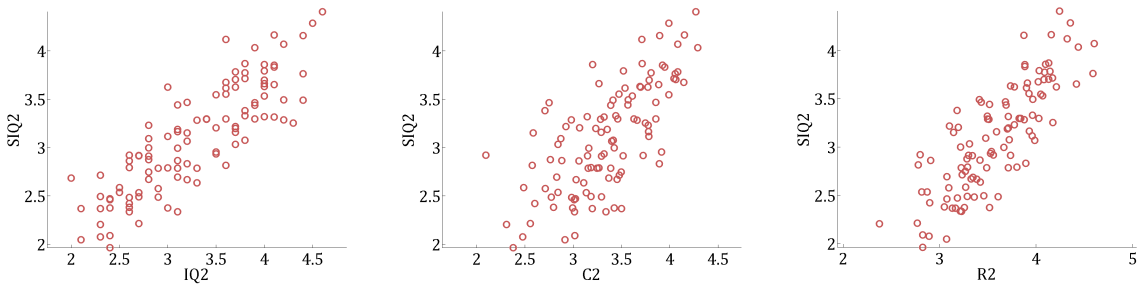


Figure 11.2: SIQ_2 plotted against IQ_2, C_2, and R_2, respectively.

Then, since the IQ_2, C_2, and R_2 mean opinion scores have been obtained during separate tests performed by different participants, no interference between the voting of these concepts

[1]Stereoscopic Image Quality Circle

has been produced and we considered them *independent* in composing the multidimensional SIQ percept.

Consequently, the model that we proposed was formulated as follows:

$$SIQ = \alpha \cdot IQ + \beta \cdot C + \gamma \cdot R + \delta \tag{11.1}$$

and the parameters fitted using *linear regression* are given in Table 11.3, along with the statistical data on their significance.

Table 11.3: Statistical data on the significance of the estimated values for the SIQ model coefficients.

	estimate	SE	t-statistic	p
δ	-0.37	0.1842	-2.0326	0.0446
α	0.48	0.0516	9.321	$\ll 0.01$
β	0.23	0.0874	2.5952	0.0108
γ	0.32	0.1095	2.912	0.0044
	RMSE: 0.239			
	Adjusted R^2: 0.817			
	F-statistic *vs.* constant model: 160, $p \ll 0.01$			

The reliability of this fitted model is confirmed by the linearity observed when plotting the estimated SIQ values against the observed SIQ_2 values (*i.e.* the MOS), as illustrated by Figure 11.3.

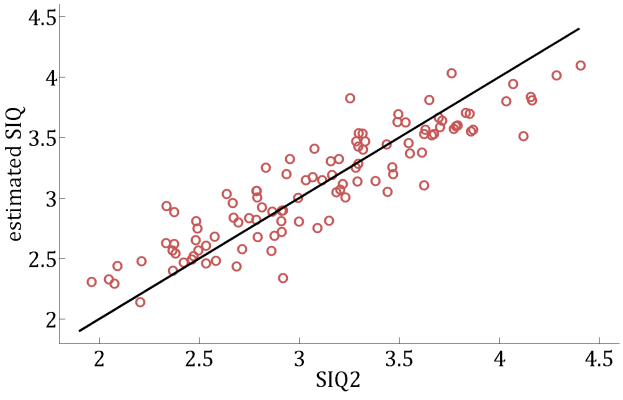


Figure 11.3: The estimated values versus the observed values for the SIQ percept.

The exact mathematical form of the model fitted is therefore:

$$SIQ = 0.48 \cdot IQ + 0.23 \cdot C + 0.32 \cdot R - 0.37. \tag{11.2}$$

This result is important in the sense that it successfully explains a part of the SIQC, the

part that is illustrated in Figure 11.4.

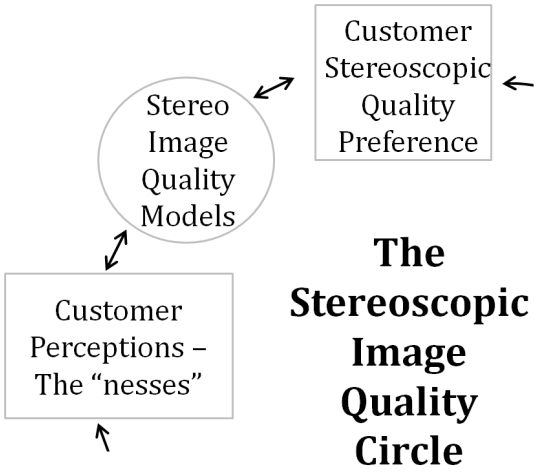


Figure 11.4: The part in the SIQC addressed by the results in Equation 11.2.

The stereoscopic image quality model proposed creates the connection between the individual percepts that observers identify while watching stereoscopic images and their overall judgment on the quality of those images, thus confirming mathematically the composition of the multidimensional concept of *stereoscopic image quality*.

11.3 Explaining the perceptual attributes with visual algorithms

Once the perceptual stereoscopic image quality model in place, the focus was on the *visual algorithms* that could be used to explain the subjective scores assigned to the *perceptual attributes* of our model, given the *physical properties* of the stereoscopic images in our datasets.

In concrete terms, this meant explaining *image quality (IQ)*, *comfort (C)*, and *realism (R)* in function of *blur*, *crossed disparities*, or *cardboard effect*.

This study was equivalent to bringing a mathematical formalization to the results of our two quantitative experiments, concentrated in Figure 9.17. Since we disposed of two different sets of data corresponding to complementary results, we performed the study separately on the data of each experiment. The findings of each of the two tests are reproduced in Figure 11.5, while the relations represented by flashes in this figure were translated into the mathematical formulas in Equations 11.3c and 11.4, by considering the observations made in Section 11.1 on the factors that were maintained constant for each of the experiments:

For Experiment 2:

$$IQ = f_{IQ}(blur), \tag{11.3a}$$
$$C = f_C(blur, crossed\_disp), \tag{11.3b}$$
$$R = f_R(blur, crossed\_disp, cdb\_eff = constant). \tag{11.3c}$$

For Experiment 3:

$$R = f_R(blur = constant, crossed\_disp = constant, cdb\_eff). \tag{11.4}$$

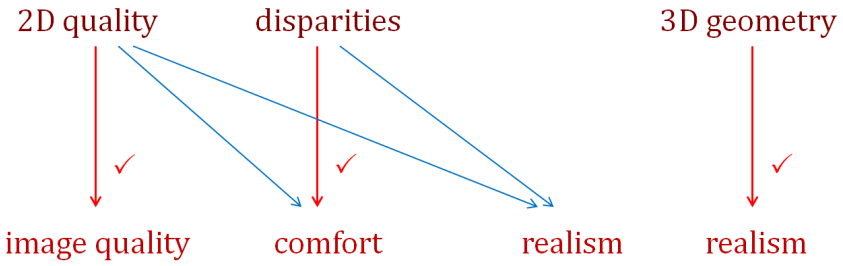


Figure 11.5: The relations between the physical image parameters and the perceptual attributes to be mathematically modeled.

Concerning the perceptual attributes that were influenced by several physical parameters, it is necessary to make the precision that the *interaction effects* were taken into account as well. The structures of our datasets, with one stereoscopic image for every possible combination of test conditions, made this possible. This meant that the influence of every parameter on the subjective ratings was recorded in the presence of variations of all the others parameters, leading to possible interactions between them. We took therefore this observation into account when searching for the best models to fit our data.

Next, for finding the appropriate mathematical models to define the four functions just presented (f_{IQ}, f_C, and f_R), we analyzed all the relations between the variables of each function and its corresponding percept. These analyses are illustrated next, separately for each perceptual attribute, along with the models to which they led.

11.3.1 Explaining the image quality

The pattern of the influence of the *blur* variations on the subjective scores on *image quality* is reproduced in Figure 11.6.

In such a case, where subjective scores are to be explained with objective measures, the recommended model for performing this mapping would be by using a logistic function [ITU 2012b]:

$$IQ = \frac{a_{IQ}}{1 + exp(b_{IQ} \cdot blur + c_{IQ})} + d_{IQ}. \tag{11.5}$$

However, despite the fact that this model approximated well the shape of our data distribution, as shown in Figure 11.7, the statistical reliability of its coefficients was not sufficient. This can be observed in Table 11.4.

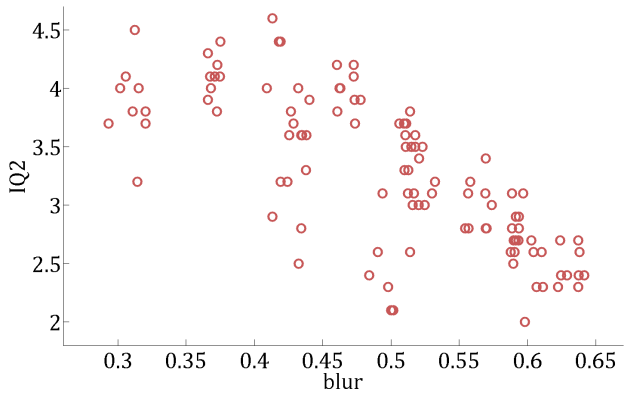


Figure 11.6: The IQ mean opinion scores plotted against the blur measures.

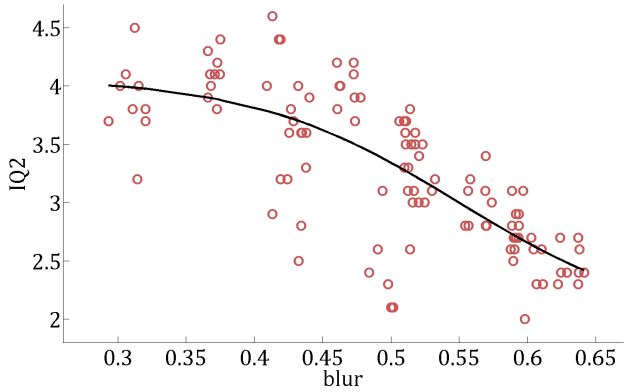


Figure 11.7: Fitting the data on IQ with a logistic function.

Table 11.4: Statistical data on the significance of the estimated values for the model coefficients when fitting the IQ data with a nonlinear function.

	estimate	SE	t-statistic	p
a_{IQ}	2.13	1.144	1.8663	0.0648
b_{IQ}	13.32	8.189	1.6261	0.1069
c_{IQ}	-7.31	3.8027	-1.9226	0.05726
d_{IQ}	1.94	0.9481	2.0451	0.04336
	RMSE: 0.429			
	Adjusted R^2: 0.562			
	F-statistic *vs.* constant model: 46.7, $p \ll 0.01$			

We could explain this behavior with the structure of our dataset, where only two variations of blur have been added to the initial images produced by our consumer camera (more details on the structure of the dataset in Appendix A.2). Initial slight variations of blur already existed in these original camera images, and this can be observed in Figure 11.6, where the values corresponding to the sharpest images are not equal, but vary slightly, and where these initial

variations are dispersed throughout the rest of the dataset, due to the artificial blur added to them. Therefore, despite the visible variability of the data in Figure 11.6, the controlled added blur was limited to only two different levels and we thus considered that a precise fit to the exact shape of the logistic function that models the human psycho-visual response could not be reached with so few controlled variations as those in our dataset.

As a consequence, other models have been tested and the one leading to the best results was the linear model of the form:

$$IQ = a_{IQ} + b_{IQ} \cdot blur, \tag{11.6}$$

with the coefficients determined by linear regression given in Table 11.5.

Table 11.5: Statistical data on the significance of the estimated values for the model coefficients when fitting the IQ data with a linear function.

	estimate	SE	t-statistic	p
a_{IQ}	5.79	0.2272	25.483	$\ll 0.01$
b_{IQ}	-5.08	0.4521	-11.237	$\ll 0.01$
	RMSE: 0.44			
	Adjusted R^2: 0.539			
	F-statistic *vs.* constant model: 126, $p \ll 0.01$			

The model that explains the image quality percept in function of the measured blur was thus:

$$IQ = 5.79 - 5.08 \cdot blur. \tag{11.7}$$

And the shape of the fitted model, superposed on the data represented in Figure 11.6, is shown if Figure 11.8.

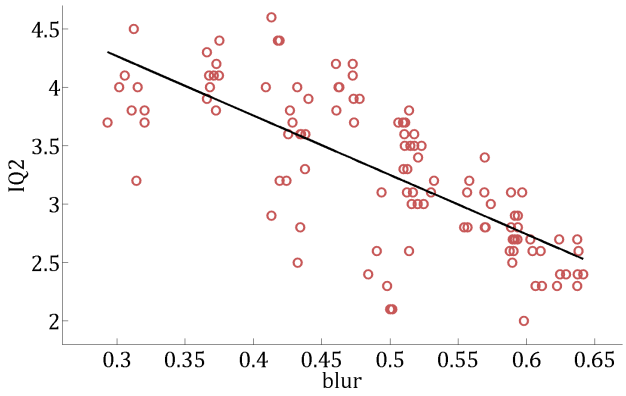


Figure 11.8: The fitted model for IQ.

11.3.2 Explaining the comfort

The *comfort* percept had proved to be the subject of a double influence, that of the level of *crossed disparities*, and also that of the *blur*. Two scatter plots are illustrated in Figure 11.9 which show how each one of these two influences could be modeled.

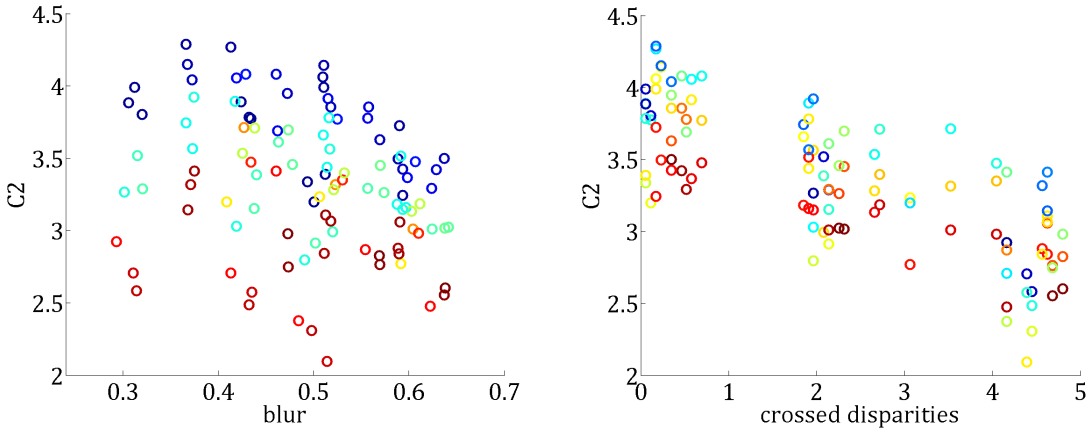


Figure 11.9: The C mean opinion scores plotted against the blur and the crossed disparities measures. The different colors in the scatter plots represent different crossed disparities in the left subfigure and different blur measures in the right subfigure.

A first modeling of the comfort percept in function of the measures on blur and crossed disparities has been performed, using a logistic function similar to that in Equation 11.5. However, we do not reproduce here the results since statistically non-reliable coefficients have been obtained in this case as well.

We thus focused on the scatter plots in Figure 11.10 in order to identify another more suitable fit.

The first plot in Figure 11.9 contains very dispersed data, however, the representation using color shows that scores of different colors, therefore corresponding to different disparities, follow the same *linear* slope. The distribution of correspondences in the second plot also has a *linear* shape, therefore the model we proposed for testing was:

$$C = a_C + b_C \cdot blur + c_C \cdot crossed\_disp + d_C \cdot blur \cdot crossed\_disp. \tag{11.8}$$

In order to better understand the *interaction effect* between the blur and the crossed disparities factors, we went back to the data of our quantitative experiments. We could thus observe in Figure 11.10 that, if representing the subjective ratings grouped in function of the three blur levels of the stereoscopic images, detailed in each group in function of the three levels of crossed disparities, the slopes in each group were similar, indicating a minimum interaction between blur and crossed disparities.

This insignificant interaction effect was confirmed by a small statistical significance of the coefficient of the interaction term, when trying to fit the model in Equation 11.8.

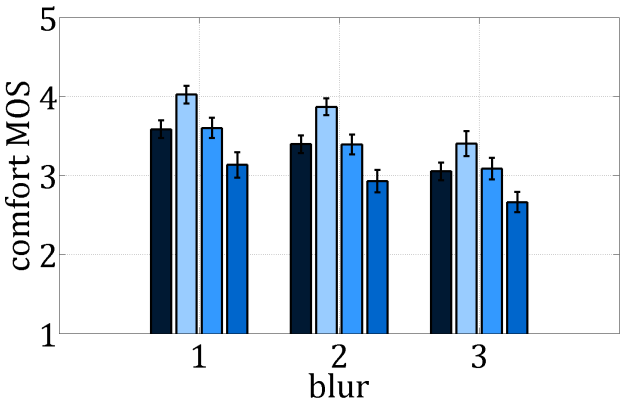


Figure 11.10: In blue, the C subjective scores grouped in function of the blur levels in the rated stereoscopic images and, for each blur group, detailed in function of the levels of crossed disparities in the images of that group; in dark blue, the mean of each blur group.

As a consequence, the interaction coefficient was not statistically validated and the model to fit became:

$$C = a_C + b_C \cdot blur + c_C \cdot crossed\_disp. \tag{11.9}$$

The coefficients obtained for this model with linear regression are given in Table 11.6 and they proved strong statistical reliability.

Table 11.6: Statistical data on the significance of the estimated values for the C model coefficients.

	estimate	SE	t-statistic	p
a_C	4.46	0.1529	29.164	$\ll 0.01$
b_C	-1.28	0.2970	-4.3191	$\ll 0.01$
C_c	-0.22	0.0169	-12.67	$\ll 0.01$
	RMSE: 0.289			
	Adjusted R^2: 0.628			
	F-statistic *vs.* constant model: 91.4, $p \ll 0.01$			

Consequently, the exact model obtained that explains the comfort percept in function of the blur and of the crossed disparities in the stereoscopic images is:

$$C = 4.46 - 1.28 \cdot blur - 0.22 \cdot crossed\_disp. \tag{11.10}$$

And the graphical representation of the fitted plane, from two slightly different angles, is in Figure 11.11, with the estimated values versus observed values representation in Figure 11.12.

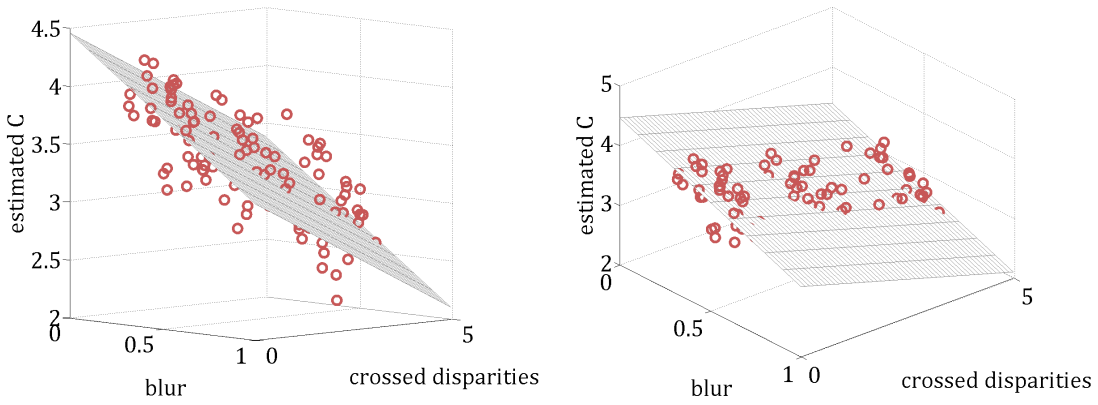


Figure 11.11: The fitted model for C.

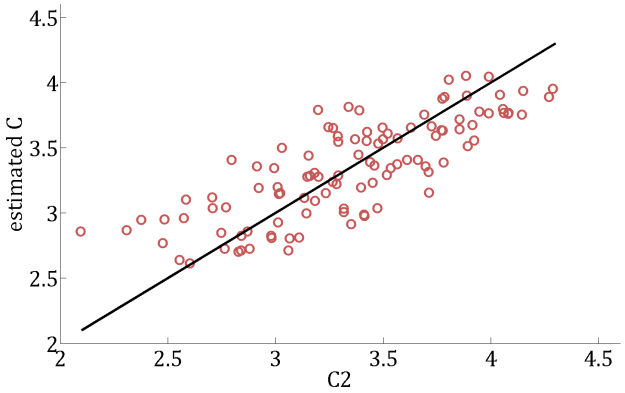


Figure 11.12: The estimated values versus the observed values for the C percept.

11.3.3 Explaining the realism

For the data collected during Experiment 2, since the *realism* was also the subject of a double influence of *blur* and *crossed disparities*, a logic similar to that of the previous section was used as well.

The relations between the subjective scores on realism and the objective values of the parameters that influence it are shown in Figure 11.13 and were used in order to decide which was the best model to fit.

The shapes of the two scatter plots also suggested a *linear* relation:

$$R = a_R + b_R \cdot blur + c_R \cdot crossed\_disp + d_R \cdot blur \cdot crossed\_disp. \tag{11.11}$$

The interaction effect between the blur and the crossed disparities seemed of significance in this case, when analyzing the graphical representation on the realism subjective scores, similar to that in Figure 11.10. This representation on the realism scores is given in Figure 11.14 and obviously shows different slopes for the scores in function of the crossed disparities, when the

blur groups are different.

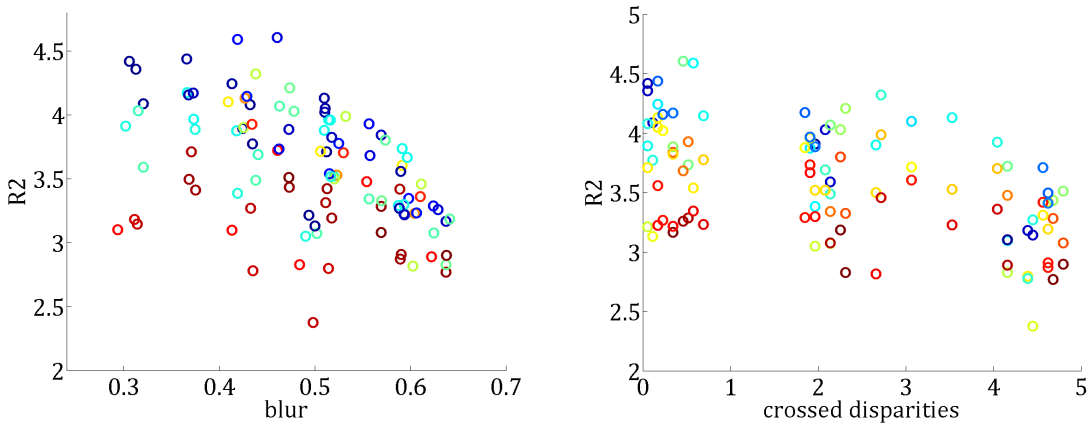


Figure 11.13: The R mean opinion scores plotted against the blur and the crossed disparities measures.

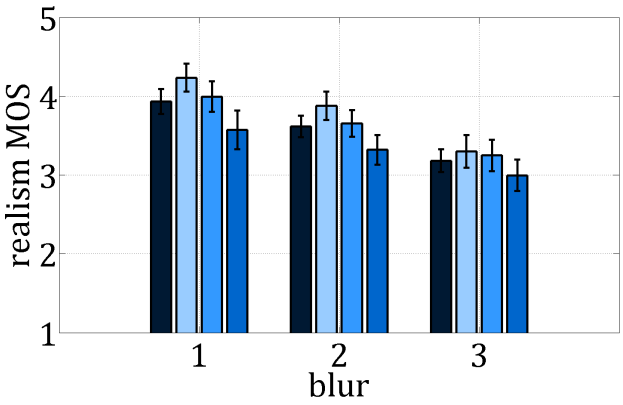


Figure 11.14: In blue, the R subjective scores grouped in function of the blur levels in the rated stereoscopic images and, for each blur group, detailed in function of the levels of crossed disparities in the images of that group; in dark blue, the mean of each blur group.

The initial model hypothesized in Equation 11.11 was thus maintained and led to a best fit for our data for the linear regression parameters shown in Table 11.7.

The exact model that explained the *realism* percept in function of the blur and of the crossed disparities in the stereoscopic images thus became:

$$R = 5.6 - 3.53 \cdot blur - 0.4 \cdot crossed\_disp + 0.55 \cdot blur \cdot crossed\_disp. \tag{11.12}$$

And in Figure 11.15 can be seen the graphical representation of the estimated values versus the original observed values.

For the data collected during Experiment 3, however, a unique influence was tested, that of the *cardboard effect* on the *realism* percept.

Table 11.7: Statistical data on the significance of the estimated values for the R model coefficients.

	estimate	SE	t-statistic	p
a_R	5.61	0.2805	20.012	$\ll 0.01$
b_R	-3.53	0.5638	-6.2594	$\ll 0.01$
C_R	-0.40	0.0997	-4.0337	$\ll 0.01$
d_R	0.55	0.1998	2.7402	$\ll 0.01$
	RMSE: 0.327			
	Adjusted R^2: 0.491			
	F-statistic *vs.* constant model: 35.4, $p \ll 0.01$			

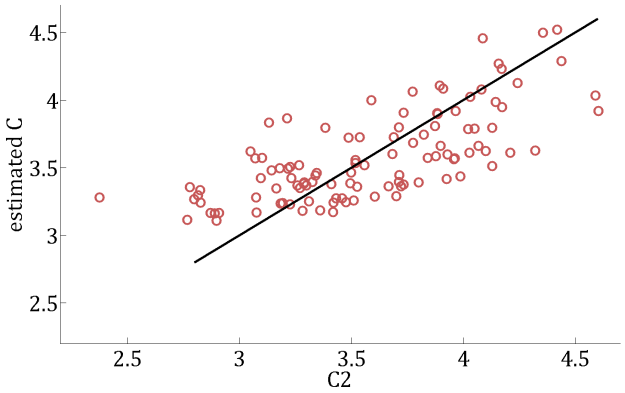


Figure 11.15: The fitted model for R in function of the blur and the disparities.

Because of the limited variation in our data, this influence was appreciated as *linear*. This is obvious in the graphical representation in Figure 11.16, which also shows that the intermediary data is missing, as mentioned in Section 11.1.

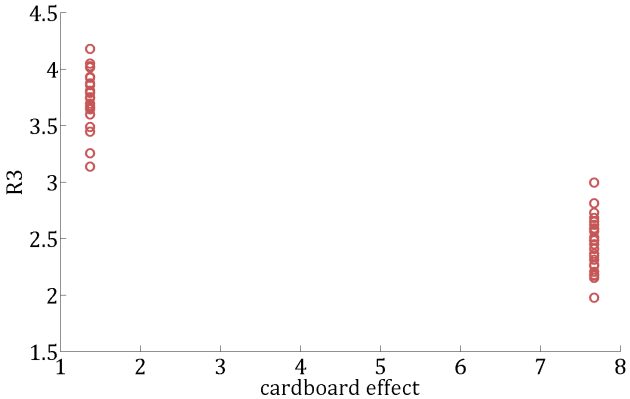


Figure 11.16: The R mean opinion scores plotted against the cardboard effect measures.

Therefore, the following formula was used for estimating realism in function of the cardboard

effect:

$$R = e_R + f_R \cdot cdb\_eff. \tag{11.13}$$

The coefficients in this formulas were estimated by linear regression to those in Table 11.8.

Table 11.8: Statistical data on the significance of the estimated values for the R model coefficients.

	estimate	SE	t-statistic	p
e_R	4.01	0.0611	65.631	$\ll 0.01$
f_R	-0.2	0.0110	-18.363	$\ll 0.01$
	RMSE: 0.242			
	Adjusted R^2: 0.877			
	F-statistic *vs.* constant model: 337, $p \ll 0.01$			

And the precise formula associating the realism to the strength of the cardboard effect is:

$$R = 4.01 - 0.2 \cdot cdb\_eff. \tag{11.14}$$

Figure 11.17 shows how this fit is appropriate for our data.

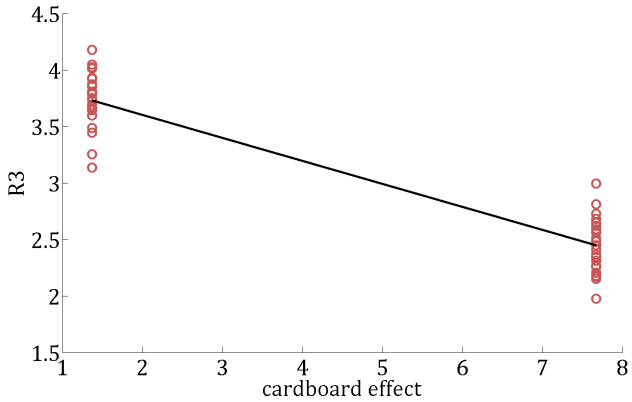


Figure 11.17: The fitted model for R in function of the cardboard effect.

We add the observation that the precision of this particular result on realism from Experiment 3 is limited due to the limited variation of the objective scores that we determined on the cardboard effect.

Now that the realism percept was explained for each one of the two experimental contexts, by equations 11.12 and 11.14, the general formula that would include the two models could be determined. The form of the model searched was determined from the relations illustrated in Figure 9.17 and was formulated as:

$$R = a \cdot blur + b \cdot crossed\_disp + c \cdot cdb\_eff + d \cdot blur \cdot crossed\_disp + e \cdot blur \cdot cdb\_eff + f \cdot crossed\_disp \cdot cdb\_eff. \quad (11.15)$$

However, adapted to our experimental data, where no interactions between the blur and the cardboard effect or between the disparities and the cardboard effect were identified, the model could be simplified to:

$$R = b_R \cdot blur + c_R \cdot crossed\_disp + d_R \cdot blur \cdot crossed\_disp + f_R \cdot cdb\_eff. \quad (11.16)$$

And, since, in Equation 11.12, the constant term stood for the constant cdb_eff term and, in Equation 11.14, the constant term stood for the constant $blur$ and $crossed\_disp$ terms, the formula on realism corresponded actually to the following explicit form:

$$R = -3.53 \cdot blur - 0.4 \cdot crossed\_disp + 0.55 \cdot blur \cdot crossed\_disp - 0.2 \cdot cdb\_eff. \quad (11.17)$$

11.3.4 Discussion

The results relating the perceptual attributes assessed by the observers with physical stereoscopic image parameters represented another important link in the SIQC. They give possible answers to the part of the Circle that is represented in Figure 11.18.

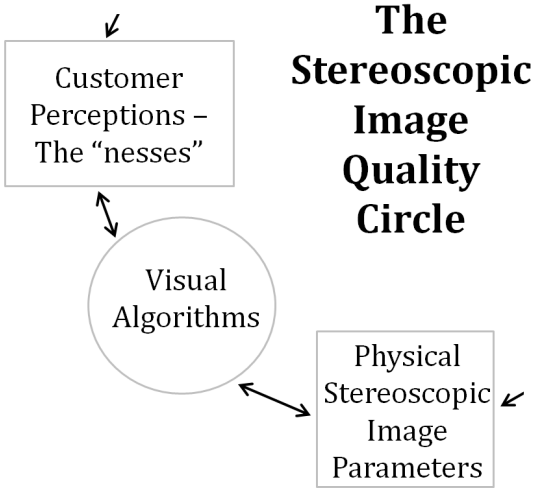


Figure 11.18: The part in the SIQC addressed by the results in Equations 11.7, 11.10, and 11.17.

We must make here the observation that these identified and quantified correlations are of course not exhaustive. By choosing blur, crossed disparities, and cardboard effect as factors influencing image quality, comfort, and realism, we had the sole purpose of being able to control

several basic parameters in our datasets in order to observe how each of the perceptual attributes identified during our exploratory study vary accordingly.

However, numerous other physical image parameters could be studied in a similar fashion and their correlation to the perceptual attributes could enrich our model. The reason why we considered only three such parameters was exclusively practical, since testing the influence of more physical parameters on the subjective percepts would have exponentially increased the duration of our experiments, already very time consuming in the form we implemented them.

11.4 Explaining the stereoscopic quality from physical image parameters

The ultimate practical purpose in the applications concerned by the assessment of the stereoscopic image quality is to be able to *standardize the procedure* such that generalized algorithms could determine an overall stereoscopic image quality rating in function of the measurable physical properties of the stereoscopic images under test.

We consider that this goal is ambitious and that implementing such a universal method would need very numerous experimental studies similar to ours before mastering all the intricate relations among the components of the SIQC.

However, we present in this section a first step towards such a method, in which the stereoscopic image quality is explained in function of the physical image parameters that varied in our study.

In the preceding sections, we showed, in Equations 11.7, 11.10, and 11.17, the mathematical formulas determined that explain the relations between the three physical parameters considered (blur, crossed disparities, and cardboard effect) and the three perceptual attributes identified (image quality, comfort, and realism). We reproduce them grouped here:

$$IQ = 5.79 - 5.08 \cdot blur, \tag{11.18a}$$

$$C = 4.46 - 1.28 \cdot blur - 0.21 \cdot crossed\_disp, \tag{11.18b}$$

$$R = -3.53 \cdot blur - 0.4 \cdot crossed\_disp + 0.55 \cdot blur \cdot crossed\_disp - 0.2 \cdot cdb\_eff. \tag{11.18c}$$

We also proposed previously a stereoscopic image quality model that would integrate the same three perceptual attributes into the overall percept on the stereoscopic image quality. This model is also reproduced here:

$$SIQ = 0.48 \cdot IQ + 0.23 \cdot C + 0.32 \cdot R - 0.37. \tag{11.19}$$

Given these mathematical relations, their simple integration led to the formula that generates a stereoscopic image quality rating directly from the physical properties of the stereoscopic image. This implied formula is:

$$SIQ = 3.42 - 3.86 \cdot blur - 0.08 \cdot crossed\_disp - 0.06 \cdot cdb\_eff + 0.17 \cdot blur \cdot crossed\_disp. \quad (11.20)$$

The great advantage of this formula is that it allows simulating the human judgment on the stereoscopic image quality only by integrating objective scores computed directly from the images under test. It can thus be implemented as an algorithm and allows avoiding the time-consuming implementation of subjective experiments with human observers.

This result is equivalent to a shortcut in the SIQC, that connects directly the physical stereoscopic image parameters to the customer stereoscopic image quality rating, as shown in Figure 11.19.

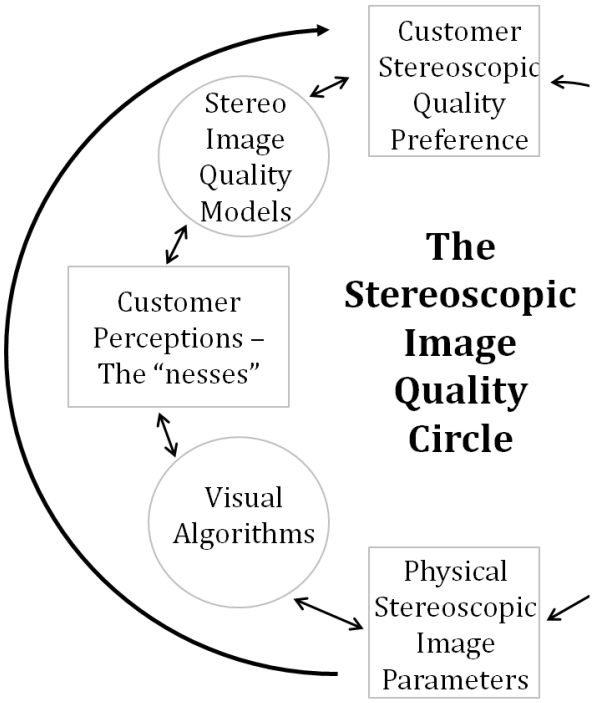


Figure 11.19: A shortcut in the SIQC.

Conclusions

This chapter concludes our investigation on the stereoscopic image quality assessment by assembling into one large picture the pieces of the puzzle collected one by one in our experimental studies or in our computations on the stereoscopic datasets created.

First, we could propose a stereoscopic image quality model that explained the multidimensional concept of *stereoscopic image quality* in function of three independent perceptual attributes: *image quality*, *comfort*, and *realism*. This corresponded to a modeling of the perceptual dimensions in the SIQC.

Next, we could determine for each one of these perceptual attributes the mathematical structures that defined them in function of three basic *physical stereoscopic image parameters*: *blur*, *crossed disparities*, and *cardboard effect*. This represented the connection between the perceptual (or subjective) and the measurable (or objective) dimensions in the SIQC.

And last, following the previously computed mathematical relations, we could derive the overall percept of *stereoscopic image quality* directly from the set of three *physical parameters*, leading to a formula exclusively based on values objectively measurable.

The findings are valuable since, based on a precise adapted formalism, *i.e.* the Stereoscopic Image Quality Circle, we could build, step by step, a method for *automatically evaluating stereoscopic images* that is *in agreement with the human judgments*.

Conclusions and Perspectives

Conclusions

As shown in the state-of-the art part of this manuscript, considerable improvements have been made to date related to stereoscopic systems. Impressive technologies already exist, the human binocular perception is better understood, promising algorithms are being implemented and tested and, in general, huge research efforts are directed towards improving every aspect related to the quality of stereoscopic data.

Despite this, it is however obvious that fully understanding the mechanisms of perceived stereoscopic quality is still a great challenge, as well as finding the ideal technique for an accurate quality evaluation, suitable in a wide range of situations.

In this context, the present work stands as an attempt of bringing new insights on the mechanisms of human perception and judgment of the stereoscopic image quality and proposes a method for assessing it.

Across the eleven chapters of this manuscript, we were thus able to first synthesize the complexity of the processes that lie behind the stereoscopic perception and behind the production chain of stereoscopic systems and to offer a broad view on the manner in which the perceived quality of this type of content can be evaluated.

Then, we directed our efforts towards better understanding what stereoscopic image quality means for an observer, through the definition of a framework, in the form of the *Stereoscopic Image Quality Circle (SIQC)*, through the setup of a complete *stereoscopic system* in our laboratory in order to perform a series of tests, through the creation of the three *MOOV3D – GIPSA Stereoscopic Datasets*, and through several *subjective experiments* rich in results.

Following our studies, we disposed of a large set of experimental data, which allowed us to elaborate a series of theoretical propositions on the perceived stereoscopic image quality, that we synthesized in Figure 11.20.

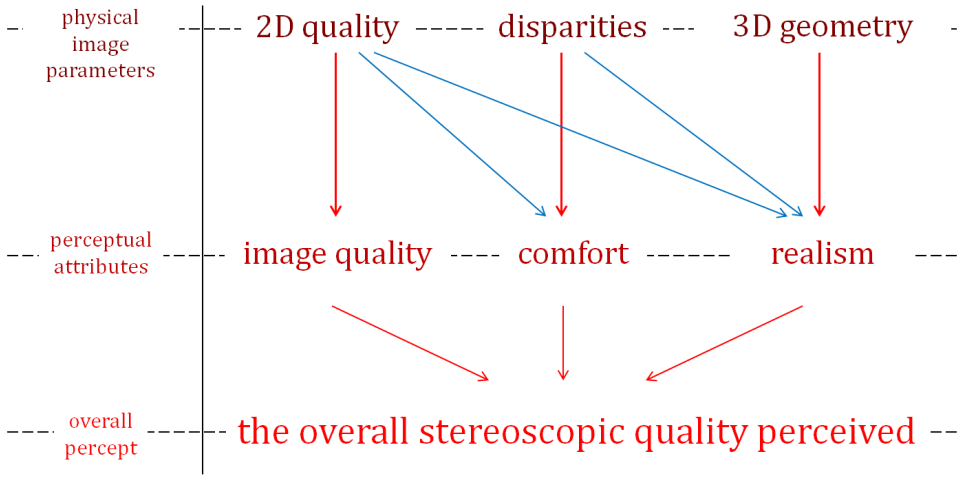


Figure 11.20: A synthesis of our results.

Thus, we proposed a *stereoscopic image quality model*, in order to explain the multidimensional concept of *stereoscopic image quality* through the means of individual *perceptual attributes*. We also used *visual algorithms* in order to explain each of the perceptual attributes of our model through the *physical parameters* of the stereoscopic images. And, finally, we could derive a *mathematical definition* of the overall percept of *stereoscopic image quality* directly based on the physical properties of the stereoscopic images. Such a definition would allow the algorithmic computation of a stereoscopic quality rating, without the need of time-consuming experimental studies.

We could observe during our subjective tests that, intuitively, people tend to judge the *stereoscopic image quality* as depending the most on the *image quality* of the composing 2D views. However, guided by clear explanations on the complexity of the stereoscopic content, even naïve observers manage to identify other factors as important in the global stereoscopic quality percept, *i.e. crossed disparities* and geometric distortions like the *cardboard effect*.

Also, the experiments we performed confirmed that unambiguously explaining the concept to rate to the participants is of extreme importance, since the public is not yet familiar with the specific vocabulary referring to stereoscopic images.

Moreover, we could have an insight on the variations of *visual comfort* when the conditions of visualization change or when the observer performs different tasks. The concentration needed for assessing precise attributes in the stereoscopic images watched can increase the visual fatigue, in tight relation with the complexity of the task. In addition, the visual fatigue was found to be stronger after visualization on a *head mounted displays* relative to a high-definition *3D TV screen*.

We consider all these findings as a small step in the elaborate undertaking towards fully explaining the human perception and judgment of the stereoscopic image quality. However, the fact that our experimental protocols allowed us to reach such elaborate and precise results encourages us to envisage a series of *perspectives* that could enrich our findings.

Perspectives

The first perspective work that we envisage is a more elaborated analysis on the *mathematical framework* proposed. We consider the results presented in Chapter 11 only a starting point. On this basis, many other mathematical models could be tested and combined into a global view on the assessment of the stereoscopic image quality in function of the physical parameters of the images. Such a mathematical formulation should also be validated by a series of new quantitative studies on different datasets and compared to other existing models.

Second, we think that reproducing our experimental protocols by considering a series of *other impairments* in the stereoscopic stimuli would lead to a broader, more complex, and more precise view on the modeling of the human judgment on stereoscopic quality.

Third, since we did not have the occasion to assess the *immersion* perceptual attribute during the quantitative experiments that we performed, a future experimental study should be implemented in order to test whether this attribute could be included as well in the stereoscopic image quality model proposed. Such a study would be even more interesting if performed on

several different stereoscopic display devices and in different contexts, since our hypothesis is that the visualization conditions would influence the most on its variations.

This third trail naturally leads us to a fourth perspective plan, that of performing more experimental studies on the *prototype 3D glasses* available in the MOOV3D project. Not all of our studies on these glasses were materialized into precise results, due to a considerable variation in the appreciations of the observers. Therefore, new exploratory studies using these glasses as a stereoscopic display would surely be an efficient method of identifying the source of these variations and of translating it in better designed quantitative experiments.

As a fifth perspective for our work, we envisage an interrogation on the perceptual attribute that we called *realism*. Despite the fact that the realism attribute, as defined by us, was influenced in our studies by all the three basic physical attributes considered in our datasets, we consider that we cannot put the equal sign between realism and stereoscopic image quality. We could have, for example, stereoscopic images considered as realistic, but which would generate discomfort, therefore the overall percept of stereoscopic quality would be deteriorated, despite the high level of realism. However, we think that an in-depth reflection on the subject and new exploratory studies might lead to another definition of this dimension or to another notion to refer to it. The phrase *realism of the rendering* might be a notion to test. It would relate this concept more to the aspects related to the 3D geometry and to the effect of the geometrical distortions and less to aspects related to the image quality of the stereoscopic images. But this is just a hypothesis and only elaborate testing could give us more clues on the best path to follow in improving the approach on this perceptual attribute.

Since we were often limited in our approach by the technical imprecision of the consumer camera we used, it would surely be interesting to set as a sixth perspective project to perform similar studies by using another *stereoscopic capture system*, composed of two high-end 2D cameras, that could be manipulated with more precision and in more diverse ways.

The seventh and final perspective work considered is to extend the insights obtained on the quality of stereoscopic still images to *stereoscopic videos*, by studying the differences in perception between the two and the manner in which these differences could be integrated in the framework that we were able to build.

By taking into account all our dense results and also all the promising premises for new research directions, we are convinced that the future findings into the human assessment of the stereoscopic image quality will transform the ultimate goal of defining the perceived stereoscopic quality from ‘almost impossible’ to ‘very complex, but feasible’.

List of Figures

List of Tables

Table of Abbreviations and Acronyms

HVS	human visual system
IPD	inter-pupillary distance
VM circle	Vieth-Müller circle
DOF	depth of focus / depth of field
ZCSBV	zone of clear, single binocular vision
ZoC	zone of comfort
ROI	region of interest
LCD	liquid-crystal display
CCD	charge-coupled device
WVGA	wide video graphics array
HDMI	high-definition multimedia interface
DCT	discrete cosine transform
HDTV	high-definition television
MOS	mean opinion score
DMOS	difference mean opinion scores
SS	single stimulus
SSCQE	single stimulus continuous quality evaluation
SDSCE	simultaneous double stimulus for continuous evaluation
ITU	International Telecommunication Union
ACR	absolute category rating
DSCQS	double stimulus continuous quality scale
SC	stimulus comparison
PC	paired comparison
IBQ	Interpretation Based Quality
CoU	context of use
PSNR	peak signal-to-noise ratio
SSIM	structural similarity index

VQM	Video Quality Metric
UQI	Universal Quality Index
RRIQA	Reduced Reference Image Quality Assessment
PQM	Perceptual Quality Metric
SSA	Stereo Sense Assessment
SBLC	Stereo Band Limited Contrast
CWT	complex wavelet transform
SSQ	Simulator Sickness Questionnaire
CRT	cathode ray tube
HMD	head-mounted display
IQC	Image Quality Circle
SIQC	Stereoscopic Image Quality Circle
MOOV3D	MObile Original Video 3D
N	nausea
O	oculomotor
D	disorientation
TS	total score
SIFT	scale-invariant feature transform
MPO	Multi-Picture Format
QoE	Quality of Experience

Bibliography

[Aflaki 2010a] P Aflaki, M Hannuksela, J Häkkinen, P Lindroos et M Gabbouj. *Impact of downsampling ratio in mixed-resolution stereoscopic video.* In 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010, pages 1–4, Tampere, 2010. (Referenced on page 52.)

[Aflaki 2010b] P Aflaki, M Hannuksela, J Hakkinen, P Lindroos et M Gabbouj. *Subjective study on compressed asymmetric stereoscopic video.* In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 4021–4024, Hong Kong, 2010. (Referenced on page 34.)

[Benoit 2008] A Benoit, P Le Callet, P Campisi et R Cousseau. *Quality Assessment of Stereoscopic Images.* EURASIP Journal on Image and Video Processing, vol. 2008, no. 1, pages 1–13, 2008. (Referenced on page 81.)

[Bensalma 2010] R Bensalma et C Larabi. *A stereoscopic quality metric based on binocular perception.* In Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on, numéro Isspa, pages 41–44, Kuala Lumpur, 2010. (Referenced on page 82.)

[Blakemore 1970] Colin Blakemore. *The range and scope of binocular depth discrimination in man.* The Journal of Physiology, pages 599–622, 1970. (Referenced on page 17.)

[Boev 2008] Atanas Boev, D Hollosi et A Gotchev. *Classification of stereoscopic artefacts.* Rapport technique, Mobile3DTV Project, 2008. (Referenced on pages 34, 53, 54, 57, 58, 59, 60, 61 et 62.)

[Boev 2009a] Atanas Boev, A Gotchev et K Egiazarian. *Stereoscopic artifacts on portable autostereoscopic displays: what matters?* In The International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-09, Scottsdale, Arizona, US, 2009. (Referenced on page 42.)

[Boev 2009b] Atanas Boev, Danilo Hollosi, Atanas Gotchev et Karen Egiazarian. *Classification and simulation of stereoscopic artifacts in mobile 3DTV content.* In Andrew J. Woods, Nicolas S. Holliman et John O. Merritt, editeurs, Stereoscopic Displays and Applications XX, San Jose, CA, USA, February 2009. (Referenced on pages 53 et 54.)

[Boev 2010] Atanas Boev, M Poikela, A Gotchev et A Aksay. *Modelling of the Stereoscopic HVS.* Rapport technique, Mobile3DTV, 2010. (Referenced on pages 23, 25, 26, 27 et 28.)

[Borel 2013] Thierry Borel et Didier Doyen. *3D Display Technologies.* In Frédéric Dufaux, Béatrice Pesquet-Popescu et Marco Cagnazzo, editeurs, Emerging Technologies for 3D Video - Creation, Coding, Transmission, and Rendering, pages 295–312. Wiley, 2013. (Referenced on pages 37, 38, 39 et 42.)

[Bouguet 2010] Jean-Yves Bouguet. *Camera Calibration Toolbox for Matlab – http://www.vision.caltech.edu/bouguetj/calib_doc/*, 2010. (Referenced on page 100.)

[Brewster 1856] Sir David Brewster. The stereoscope: its history, theory, and construction. Hastings-on-Hudson, N.Y., Morgan & Morgan, 1856. (Referenced on page 10.)

[Campisi 2007] P Campisi, P Le Callet et E Marini. *Stereoscopic Images Quality Assessment.* In European Signal Processing Conference (EUSIPCO) 2007, Poznan, Poland, 2007. (Referenced on page 80.)

[Cancellaro 2010] M Cancellaro, V Palma et A Neri. *Stereo Video Artifacts Introduced by a Distributed Coding Approach.* In 5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), 2010. (Referenced on pages 53 et 59.)

[Carnec 2003] M. Carnec, Patrick Le Callet et D. Barba. *An image quality assessment method based on perception of structural information.* In International Conference on Image Processing, pages III – 185–8 vol.2, 2003. (Referenced on page 80.)

[Chanquoy 2005] Lucile Chanquoy. Statistiques appliquées à la psychologie et aux sciences humaines et sociales. Hachette, Paris, 2005. (Referenced on pages 230, 231 et 232.)

[Chen 2010] W Chen, J Fournier, M Barkowsky et P Le Callet. *New Requirements of Subjective Video Quality Assessment Methodologies for 3DTV.* In Fifth International Workshop on Video Processing and Quality Metrics (VPQM), volume 2010, Scottsdale, US, 2010. (Referenced on page 62.)

[Chen 2012a] W Chen. *Multidimensional characterization of quality of experience of stereoscopic 3D TV.* PhD thesis, IRCCyN, 2012. (Referenced on page 52.)

[Chen 2012b] Wei Chen, Jérôme Fournier, Marcus Barkowsky et Patrick Le Callet. *Quality of experience model for 3DTV.* In Andrew J. Woods, Nicolas S. Holliman et Gregg E. Favalora, editeurs, Stereoscopic Displays and Applications XXIII, volume 8288, February 2012. (Referenced on pages 83 et 138.)

[Chen 2013] Ming-Jun Chen, Lawrence K Cormack et Alan C Bovik. *No-reference quality assessment of natural stereopairs.* IEEE Transactions on Image Processing, vol. 22, no. 9, pages 3379–91, September 2013. (Referenced on page 82.)

[Crete 2007] F Crete, T Dolmiere, P Ladret et M Nicolas. *The Blur Effect : Perception and Estimation with a New No-Reference Perceptual Blur Metric.* In SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging, volume 6492, 2007. (Referenced on pages 225 et 243.)

[Cutting 1995] James E. Cutting et Peter M. Vishton. *Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth.* Handbook of perception and cognition, vol. 5, pages 69–117, 1995. (Referenced on pages 25, 26 et 27.)

[Devernay 2010] F Devernay et P Beardsley. Stereoscopic Cinema, volume 5 of *Geometry and Computing.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. (Referenced on pages 12, 14, 26, 33, 60 et 63.)

[Dodgson 2004] Neil A Dodgson. *Variation and extrema of human interpupillary distance.* In Andrew Woods, J O Merrit, S A Benton et M T Bolas, editeurs, SPIE Stereoscopic Displays and Virtual Reality Systems X1, volume 5291, pages 36–46, San Jose, CA, USA, 2004. (Referenced on page 16.)

[Eerola 2008] T Eerola, J K Kamarainen, T Leisti, R Halonen, L Lensu, H Kalviainen, G Nyman et P Oittinen. *Is there hope for predicting human visual quality experience?* In 2008 IEEE International Conference on Systems Man and Cybernetics, pages 725–732, Singapore, 2008. IEEE. (Referenced on page 83.)

[Emoto 2005] M. Emoto, T. Niida et F. Okano. *Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television.* Journal of Display Technologies, vol. 1, no. 2, pages 328–340, 2005. (Referenced on pages 24 et 65.)

[Endrikhovski 2005] Serguei Endrikhovski, Elaine Jin, Michael E. Miller et Robert W. Ford. *Predicting individual fusional range from optometric data.* In Andrew J. Woods, Mark T. Bolas, John O. Merritt et Ian E. McDowall, editeurs, Proceedings of SPIE-IS&T Electronic Imaging, volume 5664, pages 36–47, San Jose, CA, 2005. (Referenced on page 20.)

[Engeldrum 1989] Peter G Engeldrum. *Measuring customer perception of print quality.* In IS&T 42nd Annual Meeting, numéro March, pages 161–164, 1989. (Referenced on page 85.)

[Engeldrum 1999] Peter G Engeldrum. *Image Quality Modeling : Where Are We ?* In IS&T's PICS Conference, pages 251–255, 1999. (Referenced on pages 67 et 85.)

[Engeldrum 2001] Peter G Engeldrum. *Psychometric Scaling : Avoiding the Pitfalls and Hazards.* In PICS: Image Processing, Image Quality, Image Capture Systems Conference, pages 101–107, Montreal, Canada, 2001. (Referenced on pages 67, 68 et 74.)

[Engeldrum 2004a] Peter G Engeldrum. *A Short Image Quality Model Taxonomy.* Journal of Imaging Science and Technology, vol. 48, no. 2, pages 160–165, 2004. (Referenced on page 50.)

[Engeldrum 2004b] Peter G Engeldrum. *A Theory of Image Quality : The Image Quality Circle.* Journal of Imaging Science and Technology, vol. 48, no. 5, pages 446–456, 2004. (Referenced on pages 67, 85 et 95.)

[Fujifilm FinePix REAL 3D W3 2013] Fujifilm FinePix REAL 3D W3. *http://www.fujifilm.com/products/3d/camera/finepix_real3dw3/*, 2013. (Referenced on page 99.)

[Gorley 2008] Paul Gorley et Nick Holliman. *Stereoscopic image quality metrics and compression.* SPIE Stereoscopic Displays and Applications, vol. 6803, no. January, pages 680305–680305–12, 2008. (Referenced on pages 22, 80 et 81.)

[Gorley 2010] Paul Gorley et Nicolas Holliman. *Investigating symmetric and asymmetric stereoscopic compression using the PSNR image quality metric.* In Information Sciences and Systems (CISS), 2010 44th Annual Conference on, pages 1–6, 2010. (Referenced on pages 34 et 35.)

[Gotchev 2008] Atanas Gotchev, Satu Jumisko-Pyykkö, Atanas Boev et Dominik Strohmeier. *Mobile 3DTV System: Quality and User Perspective.* In European Workshop of Media Delivery (EUMOB'08), Oulu, Finland, 2008. ICST. (Referenced on page 76.)

[Gunnewiek 2010] R.K. Gunnewiek et P Vandewalle. *How to display 3D content realistically.* In Proc. Int. Workshop Video Process. Quality Metrics Consum. Electron.(VPQM), 2010. (Referenced on pages 43 et 45.)

[Häkkinen 2002] Jukka Häkkinen, Tero Vuori, Monika Puhakka, B Postural et A Participants. *Postural stability and sickness symptoms after HMD use.* In IEEE International Conference on Systems, Man and Cybernetics, pages 147–152, Yasmine Hammamet, Tunisia, 2002. (Referenced on page 78.)

[Häkkinen 2006a] J Häkkinen, M Polonen, J Takatalo et G Nyman. *Simulator Sickness in Virtual Display Gaming : A Comparison of Stereoscopic and Non-stereoscopic.* In Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, pages 227–229, Helsinki, Finland, 2006. (Referenced on pages 65 et 78.)

[Häkkinen 2006b] J Häkkinen, J Takatalo, J Komulainen, H Särkelä, J Havukumpu et G Nyman. *Simulator sickness symptoms in virtual display gaming.* In Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services, pages 227–230, 2006. (Referenced on page 78.)

[Hanhart 2013] Philippe Hanhart, Francesca De Simone, Martin Rerabek et Touradj Ebrahimi. *3D Video Quality Assessment.* In Frédéric Dufaux, Béatrice Pesquet-Popescu et Marco Cagnazzo, editeurs, Emerging Technologies for 3D Video - Creation, Coding, Transmission, and Rendering, pages 377–391. 2013. (Referenced on pages 53, 57, 62, 68 et 75.)

[Hewage 2008] C Hewage, S Worrall, S Dogan et A Kondoz. *Prediction of stereoscopic video quality using objective quality models of 2-D video.* Electronics Letters, vol. 44, no. 16, pages 963 – 965, 2008. (Referenced on page 79.)

[Holliman 2006] N Holliman. *3D Display Systems.* In J Dakin et R Brown, editeurs, Handbook of Optoelectronics, pages 1067–1100. Taylor & Francis, New York and London, 2006. (Referenced on pages 25, 26, 39, 40, 41, 42, 43 et 60.)

[Howard 1995] IP Howard et BJ Rogers. Binocular vision and stereopsis. Oxford Psychology Series, 1995. (Referenced on page 28.)

[Howarth 1997] P. A. Howarth et P. J. Costello. *The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system.* Displays, vol. 18, no. 2, pages 107–116, December 1997. (Referenced on pages 78, 102 et 164.)

[IJsselsteijn 2000] W IJsselsteijn, H Ridder et J Vliegen. *Subjective evaluation of stereoscopic images: effects of camera parameters and display duration.* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 10, no. 2, pages 225–233, 2000. (Referenced on page 32.)

[IJsselsteijn 2002a] W IJsselsteijn, D Bouwhuis, J Freeman et H de Ridder. *Presence as an Experiential Metric for 3-D Display Evaluation.* In SID Symposium Digest of Technical Papers, volume 33, Boston, MA, USA, 2002. (Referenced on page 51.)

[IJsselsteijn 2002b] W IJsselsteijn, P Seuntiens et L Meesters. *State-of-the-art in human factors and quality issues of stereoscopic broadcast television.* Rapport technique, ATTEST Project, 2002. (Referenced on pages 17, 21, 23, 24, 25, 27, 33, 34, 35, 38, 55 et 62.)

[IJsselsteijn 2005] Wijnand A. IJsselsteijn, Pieter J.H. Seuntiëns et Lydia M.J. Meesters. *Human Factors of 3D Displays.* In O. Schreer, P. Kauff et T. Sikora, editeurs, 3D videocommunication: algorithms, concepts and real-time systems in human centred communication, chapitre 12, pages 219–233. John Wiley & Sons, New Jersey, 2005. (Referenced on page 28.)

[Ionescu 2009] Gelu Ionescu, Nathalie Guyader et Anne Guérin. *SoftEye Software*, 2009. (Referenced on page 112.)

[ITU-T SG9: Broadband cable and TV 2013] ITU-T SG9: Broadband cable and TV. *www.itu.int/ITU-T/studygroups/com09/*, 2013. (Referenced on page 67.)

[ITU 2008] ITU. *Recommendation ITU-T P.910 - Subjective video quality assessment methods for multimedia applications.* Rapport technique, ITU, 2008. (Referenced on page 113.)

[ITU 2012a] ITU. *Recommendation ITU-R BT.2012 - Subjective methods for the assessment of stereoscopic 3DTV systems.* Rapport technique, ITU, 2012. (Referenced on pages 67, 68 et 71.)

[ITU 2012b] ITU. *Recommendation ITU-R BT.500-13 - Methodology for the subjective assessment of the quality of television pictures.* Rapport technique, ITU, 2012. (Referenced on pages 67, 68, 69, 70, 71, 72, 73, 74, 144, 145, 146, 173, 223, 227, 228 et 229.)

[Jacobs 2002] Robert A. Jacobs. *What determines visual cue reliability?* TRENDS in Cognitive Sciences, vol. 6, no. 8, pages 345–350, August 2002. (Referenced on page 27.)

[Jin 2007] Zhao Xia Jin, Ya Jun Zhang, Xin Wang et Thomas Plocher. *Evaluating the Usability of an Auto-stereoscopic Display.* In Human-Computer Interaction. Interaction Platforms and Techniques. Lecture Notes in Computer Science, pages 605–614. 2007. (Referenced on page 65.)

[Jongerius 2012] Jerry Jongerius. *Measuring lens field of view – http://www.panohelp.com/lensfov.html*, 2012. (Referenced on pages 100 et 101.)

[Joveluro 2010] P Joveluro, H Malekmohamadi, W Fernando et A Kondoz. *Perceptual Video Quality Metric for 3D video quality assessment.* In 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010, pages 1–4, Tampere, 2010. (Referenced on page 80.)

[Julesz 1971] Bela Julesz. Foundations of cyclopean perception. University of Chicago Press, 1971. (Referenced on page 28.)

[Jumisko-Pyykko 2007a] S Jumisko-Pyykko, J Hakkinen et G Nyman. *Experienced Quality Factors - Qualitative Evaluation Approach to Audiovisual Quality.* Multimedia on Mobile Devices, SPIE, vol. 6507, 2007. (Referenced on page 75.)

[Jumisko-Pyykko 2007b] S Jumisko-Pyykko, U Reiter et C Weigel. *Produced Quality is Not Perceived Quality - A Qualitative Approach to Overall Audiovisual Quality.* In 3DTV Conference, 2007, pages 1–4, Kos Island, 2007. (Referenced on pages 75 et 76.)

[Jumisko-Pyykko 2008] S Jumisko-Pyykko et D Strohmeier. *Report on research methodologies for the experiments.* Rapport technique, Mobile3DTV, 2008. (Referenced on page 75.)

[Jumisko-Pyykko 2010a] S Jumisko-Pyykko et T Utriainen. *A Hybrid Method for Quality Evaluation in the Context of Use for Mobile (3D) Television.* Multimedia Tools and Applications, vol. 55, no. 2, pages 185–225, September 2010. (Referenced on pages 76 et 77.)

[Jumisko-Pyykko 2010b] S Jumisko-Pyykko, T Utriainen, D Strohmeier, Atanas Boev et K Kunze. *Simulator sickness – Five experiments using autostereoscopic mid-sized or*

small mobile screens. In 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, volume 3, pages 1–4, Tampere, 2010. (Referenced on pages 78 et 164.)

[Kaptein 2008] Ronald G Kaptein, Marc T M Lambooij, Wijnand A Ijsselsteijn et Ingrid Heynerickx. *Performance evaluation of 3D-TV systems.* In Susan P. Farnand et Frans Gaykema, editeurs, SPIE Image Quality and System Performance V, volume 6808, pages 1–11, San Jose, CA, USA, 2008. (Referenced on pages 51, 52 et 54.)

[Kennedy 1993] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum et Michael G. Lilienthal. *Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness.* The International Journal of Aviation Psychology, vol. 3, no. 3, pages 203–220, 1993. (Referenced on pages 78, 160, 164 et 235.)

[Knill 2003] D Knill et J Saunders. *Do humans optimally integrate stereo and texture information for judgments of surface slant?* Vision Research, vol. 43, no. 24, pages 2539–2558, November 2003. (Referenced on page 28.)

[Kooi 2004] F Kooi et A Toet. *Visual comfort of binocular and 3D displays.* Displays, vol. 25, no. 2-3, pages 99–108, August 2004. (Referenced on pages 22, 23, 61 et 65.)

[Lambooij 2007] M Lambooij, W IJsselsteijn et I Heynderickx. *Visual discomfort in stereoscopic displays: a review.* SPIE, Stereoscopic Displays and Virtual Reality Systems XIV, vol. 6490, pages 64900I.1–64900I.11, 2007. (Referenced on pages 17, 20, 23, 26, 28, 44, 52, 55, 61, 62, 63 et 77.)

[Lambooij 2009] M Lambooij, M Fortuin, W IJsselsteijn et I Heynderickx. *Measuring visual discomfort associated with 3D displays.* SPIE, vol. 7237, no. 1, pages 72370K–72370K–12, 2009. (Referenced on page 24.)

[Lambooij 2011] M Lambooij, W IJsselsteijn, D Bouwhuis et I Heynderickx. *Evaluation of Stereoscopic Images: Beyond 2D Quality.* IEEE Transactions on Broadcasting, vol. 57, no. 2, pages 432–444, June 2011. (Referenced on pages 51, 84, 85 et 138.)

[Leon 2008] G Leon, H Kalva et B Furht. *3D Video Quality Evaluation with Depth Quality Variations.* In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, pages 301 – 304, Istanbul, 2008. (Referenced on page 53.)

[Lipton 1982] Lenny Lipton. Foundations of Stereoscopic Cinema. Van Nostrand Reinhold Company, 1982. (Referenced on pages 10, 11, 13, 25 et 33.)

[Lipton 1997] Lenny Lipton. StereoGraphics Developers' Handbook. 1997. (Referenced on page 33.)

[Liu 2008] Y Liu, A Bovik et L Cormack. *Disparity statistics in natural scenes.* Journal of Vision, vol. 8, no. 11, pages 1–14, 2008. (Referenced on pages 16, 22 et 23.)

[Liu 2010] Y Liu, L Cormack et A Bovik. *Dichotomy between luminance and disparity features at binocular fixations.* Journal of Vision, vol. 10, no. 12, pages 1–17, 2010. (Referenced on page 22.)

[Lowe 1999] David G. Lowe. *Object recognition from local scale-invariant features.* In Proceedings of the IEEE International Conference on Computer Vision, pages 1150–1157 vol.2. IEEE, 1999. (Referenced on page 244.)

[Lowe 2004] David G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints.* International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, November 2004. (Referenced on page 244.)

[Lowe 2013] David G. Lowe. *Demo Software: SIFT Keypoint Detector – http://www.cs.ubc.ca/~lowe/keypoints/*, 2013. (Referenced on page 244.)

[Lu 2009] F Lu, H Wang, X Ji et G Er. *Quality assessment of 3D asymmetric view coding using spatial frequency dominance model.* In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009, numéro 60772046, pages 1–4, Potsdam, 2009. (Referenced on pages 24 et 34.)

[Lumière 1895] Louis Lumière. *Arrivée d'un train (à la Ciotat) – http://www.moma.org/collection/*, 1895. (Referenced on page 12.)

[Mao 2010] X Mao, M Yu, X Wang, G Jiang, Z Peng et J Zhou. *Stereoscopic image quality assessment model with three-component weighted structure similarity.* In Audio Language and Image Processing (ICALIP), 2010 International Conference on, pages 1175–1179, Shanghai, 2010. (Referenced on page 80.)

[Marcotorchino 1979] J Marcotorchino et P Michaud. Optimisation en analyse ordinale des données Volume 4 of Statistique et décisions économiques. Masson, Paris, 1979. (Referenced on page 232.)

[McCarthy 2010] S McCarthy. *Glossary for Video & Perceptual Quality of Stereoscopic Video.* Rapport technique, 3D@Home Consortium, MPEG Industry Forum 3DTV Working Group, 2010. (Referenced on pages 18, 20, 23, 24, 53, 58 et 62.)

[Meesters 2003] L Meesters, W IJsselsteijn et P Seuntiens. *Survey of perceptual quality issues in three-dimensional television systems.* In Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems X, volume 5006, pages 313–326, Santa Clara, CA, USA, May 2003. SPIE. (Referenced on pages 34, 35, 38, 39, 40, 41, 42, 60 et 61.)

[Meesters 2004] L Meesters, W IJsselsteijn et P Seuntiens. *A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV.* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 3, pages 381–391, 2004. (Referenced on pages 61, 75 et 77.)

[Mendiburu 2009] B Mendiburu. 3D Movie Making Stereoscopic Digital Cinema from Script to Screen. Focal Press, 2009. (Referenced on pages 14, 19, 26, 32, 33, 47, 53 et 63.)

[MicroOLED 2013] MicroOLED. *Internal report.* Rapport technique, 2013. (Referenced on page 102.)

[Nojiri 2003] Y Nojiri, H Yamanoue, A Hanazato et F Okano. *Measurement of parallax distribution, and its application to the analysis of visual comfort for stereoscopic HDTV.* In SPIE Stereoscopic Displays and Virtual Reality Systems X, volume 5006, pages 195–205, Santa Clara, CA, USA, 2003. (Referenced on page 65.)

[Nyman 2006] Göte Nyman, Jenni Radun, Tuomas Leisti, Joni Oja, Harri Ojanen, Jean-luc Olives, Tero Vuori et Jukka Häkkinen. *What do users really perceive - probing the subjective image quality.* In Luke C Cui et Yoichi Miyake, editeurs, SPIE Image Quality and System Performance III, volume 6059, pages 1–7, 2006. (Referenced on page 76.)

[Okada 2006] Y Okada, K Ukai, J Wolffsohn, B Gilmartin, A Iijima et T Bando. *Target spatial frequency determines the response to conflicting defocus- and convergence-driven accommodative stimuli.* Vision Research, vol. 46, no. 4, pages 475–84, February 2006. (Referenced on page 63.)

[Pastoor 1995] S Pastoor. *Human Factors of 3D Imaging : Results of Recent Research at Heinrich-Hertz-Institut Berlin.* Proceedings of the International Display Workshop 95, vol. 3, pages 66–72, 1995. (Referenced on pages 22 et 141.)

[Patterson 2007] Robert Patterson. *Human factors of 3-D displays.* Journal of the Society for Information Display, vol. 15, no. 11, page 861, 2007. (Referenced on pages 20, 22, 23, 24, 25 et 61.)

[Pinson 2004] M.H. Pinson et S. Wolf. *A new standardized method for objectively measuring video quality.* IEEE Transactions on Broadcasting, vol. 50, no. 3, pages 312–322, 2004. (Referenced on page 79.)

[Qin 2006] D Qin, M Takamatsu et Y Nakashima. *Disparity Limit for Binocular Fusion in Fovea.* Optical Review, vol. 13, no. 1, pages 34–38, 2006. (Referenced on pages 20, 21 et 22.)

[Rogers 1982] Brian Rogers et Maureen Graham. *Similarities between motion parallax and stereopsis in human depth perception.* Vision Research, vol. 22, pages 261–270, 1982. (Referenced on page 26.)

[Ryu 2012] Seungchul Ryu, Dong Hyun Kim et Kwanghoon Sohn. *Stereoscopic image quality metric based on binocular perception model.* In Image Processing (ICIP), 19th IEEE International Conference, numéro Diml, pages 609–612, Orlando, FL, 2012. (Referenced on page 82.)

[Sand 1992] R Sand. *3-DTV - A review of recent and current developments.* In Proceedings IEE Colloq. Stereoscopic Television, pages 1–4, 1992. (Referenced on page 13.)

[Seuntiens 2005] P Seuntiens, I Heynderickx, W Ijsselsteijn, P van den Avoort, J Berentsen, I Dalm, M Lambooij et W Oosting. *Viewing experience and naturalness of 3D images.* In SPIE 6016 - Integral Imaging Systems, volume 31, pages 0–6, Boston, MA, USA, 2005. (Referenced on page 51.)

[Seuntiens 2006a] P Seuntiens. *Visual Experience of 3D TV.* Phd thesis, Eindhoven: Technische Universiteit Eindhoven, 2006. (Referenced on pages 83, 84, 85 et 138.)

[Seuntiens 2006b] P Seuntiens, L Meesters et W IJsselsteijn. *Perceived quality of compressed stereoscopic images : effects of symmetric and asymmetric JPEG coding and camera separation.* ACM Transactions on Applied Perception, vol. 3, no. 2, pages 95 – 109, 2006. (Referenced on pages 34, 51, 52 et 65.)

[Shibata 2011] Takashi Shibata, Joohwan Kim, David M Hoffman et Martin S Banks. *The zone of comfort : Predicting visual discomfort with stereo displays.* Journal of Vision, vol. 11, no. 8, pages 1–29, 2011. (Referenced on pages 18, 52, 53, 63 et 65.)

[Smith 1997] George Smith et David A. Atchison. The Eye and Visual Optical Instruments. 1997. (Referenced on page 18.)

[Smolic 2009] A Smolic, K Mueller, P Merkle et A Vetro. *Development of a new MPEG Standard for Advanced 3D Video Applications.* In Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on, pages 400–407, Salzburg, 2009. (Referenced on page 36.)

[Sony 3D Television & Film 2010] Sony 3D Television & Film. *http://www.sony.co.uk/res/attachment/file/85/1237478817085.pdf*, 2010. (Referenced on pages 32, 37, 38 et 39.)

[Sony TX-P50VT20B 3D Plasma TV 2013] Sony TX-P50VT20B 3D Plasma TV. *http://www.panasonic.co.uk/html/en_GB/Products/TX-P50VT20B/Overview/3665336/index.html*, 2013. (Referenced on page 102.)

[Spottiswoode 1952] Raymond Spottiswoode, N L Spottiswoode et Charles Smith. *Basic Principles of the Three-Dimensional Film.* Journal of the Society of Motion Picture Television Engineers, vol. 59, pages 249–286, 1952. (Referenced on pages 12 et 33.)

[Starosta 2013] Boris Starosta. *Screensize versus depth effect – http://starosta.com/artready/screensize_VS_depth_effect/screensize_depth_.jpg*, 2013. (Referenced on pages 45 et 46.)

[Stevenson 1997] Scott B. Stevenson et Clifton M. Schor. *Human Stereo Matching is not Restricted to Epipolar Lines.* Vision Research, vol. 37, no. 19, pages 2717–2723, 1997. (Referenced on page 22.)

[Strohmeier 2010a] D Strohmeier, S Jumisko-Pyykko et K Kunze. *New, Lively, And Exciting Or Just Artificial, Straining, And Distracting: A Sensory Profiling Approach To Understand Mobile 3D Audiovisual Quality.* In The Fifth International Workshop on Video Processing and Quality Metrics (VPQM 2010), Scottsdale, Arizona, US, 2010. (Referenced on page 53.)

[Strohmeier 2010b] D Strohmeier, S Jumisko-Pyykko et U Reiter. *Profiling experienced quality factors of audiovisual 3D perception.* In Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on, pages 70–75, Trondheim, 2010. (Referenced on page 51.)

[Strohmeier 2010c] D Strohmeier et G Tech. *" Sharp , bright , three-dimensional " - Open Profiling of Quality for mobile 3DTV coding methods.* In Proceedings of SPIE, the International Society for Optical Engineering, volume 49, San Jose, California, United States, 2010. (Referenced on page 84.)

[Tam 1998] James Tam, Lew B Stelmach et Philip Corriveau. *Psychovisual aspects of viewing stereoscopic video sequences.* In SPIE Stereoscopic Displays and Applications IX, volume 3295, pages 226–235, San Jose, California, United States, 1998. (Referenced on pages 51 et 52.)

[Tam 2011] Wa James Tam, Filippo Speranza, Sumio Yano, Koichi Shimono et Hiroshi Ono. *Stereoscopic 3D-TV: Visual Comfort.* IEEE Transactions on Broadcasting, vol. 57, no. 2, pages 335–346, 2011. (Referenced on pages 52, 53, 58, 63 et 77.)

[Tominaga 2010] Toshiko Tominaga, Takanori Hayashi, Jun Okamoto et Akira Takahashi. *Performance comparisons of subjective quality assessment methods for mobile video.* In Quality of Multimedia Experience QoMEX 2010 Second International Workshop on, pages 82–87, 2010. (Referenced on page 115.)

[Torralba 2002] Antonio Torralba et Aude Oliva. *Depth Estimation from Image Structure.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pages 1226–1238, 2002. (Referenced on pages 25 et 26.)

[Ukai 2006] K Ukai. *Human Factors for Stereoscopic Images.* In Multimedia and Expo, 2006 IEEE International Conference on, pages 1697–1700, Toronto, Ont., 2006. (Referenced on pages 20 et 25.)

[Video Quality Experts Group (VQEG) 3DTV Project 2013] Video Quality Experts Group (VQEG) 3DTV Project. *http://www.its.bldrdoc.gov/vqeg/projects/3dtv/3dtv.aspx*, 2013. (Referenced on page 67.)

[Wang 2002] Zhou Wang et Alan Bovik. *A universal image quality index.* IEEE Signal Processing Letters, vol. 9, no. 3, pages 81–84, 2002. (Referenced on page 80.)

[Wang 2004] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh et Eero P Simoncelli. *Image quality assessment: from error visibility to structural similarity.* IEEE Transactions on Image Processing, vol. 13, no. 4, pages 600–12, April 2004. (Referenced on page 79.)

[Wang 2005] Zhou Wang et Eero P Simoncelli. *Reduced-reference image quality assessment using a wavelet-domain natural image statistic model.* In SPIE Human Vision and Electronic Imaging X, 2005. (Referenced on page 80.)

[Watt 2013] Simon J. Watt et Kevin J. MacKenzie. *3D Media and the Human Visual System.* In Frédéric Dufaux, Béatrice Pesquet-Popescu et Marco Cagnazzo, editeurs, Emerging Technologies for 3D Video - Creation, Coding, Transmission, and Rendering, pages 349–376. Wiley, 2013. (Referenced on pages 17, 18, 19, 22, 27, 28, 61, 62 et 64.)

[Wheatstone 1838] Charles Wheatstone. *Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision.* Philosophical Transactions of the Royal Society of London, vol. 128, pages 371–394, 1838. (Referenced on page 9.)

[Wheatstone 1852] Charles Wheatstone. *Contributions to the Physiology of Vision. Part the Second. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision.* Philosophical Transactions of the Royal Society of London, vol. 142, pages 1–17, 1852. (Referenced on page 10.)

[Woods 1993] A Woods, T Docherty et R Koch. *Image Distortions in Stereoscopic Video Systems.* In Scott S. Fisher John O. Merritt, editeur, SPIE Stereoscopic Displays and Applications IV, volume 1915, San Jose, CA, USA, 1993. (Referenced on pages 53, 101, 103 et 106.)

[Woods 2011] Andrew J. Woods. *How are crosstalk and ghosting defined in the stereoscopic literature?* In Andrew J. Woods, Nicolas S. Holliman et Neil A. Dodgson, editeurs, Stereoscopic Displays and Applications XXII, volume 7863, February 2011. (Referenced on page 61.)

[Xing 2010a] L Xing, J You, T Ebrahimi et A Perkis. *An objective metric for assessing quality of experience on stereoscopic images.* In Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on, pages 373–378, 2010. (Referenced on pages 61 et 82.)

[Xing 2010b] L Xing, J You, T Ebrahimi et A Perkis. *Estimating Quality of Experience on Stereoscopic Images.* In Intelligent Signal Processing and Communication Systems (ISPACS), 2010 International Symposium on, numéro lSPACS, pages 1–4, Chengdu, 2010. (Referenced on page 82.)

[Yamanoue 2000] H Yamanoue, M Okui et I Yuyama. *A study on the relationship between shooting conditions and cardboard effect of stereoscopic images.* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 10, no. 3, pages 411–416, 2000. (Referenced on page 150.)

[Yang 2009] J Yang, C Hou, Y Zhou, Z Zhang et J Guo. *Objective quality assessment method of stereo images.* In 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009, pages 1–4, Potsdam, 2009. (Referenced on page 80.)

[Yano 2002] Sumio Yano, Shinji Ide, Tetsuo Mitsuhashi et Hal Thwaites. *A study of visual fatigue and visual comfort for 3D HDTVHDTV images.* Displays, vol. 23, pages 191–201, 2002. (Referenced on page 65.)

[Yasakethu 2009] S Yasakethu, W Fernando, B Kamolrat et A Kondoz. *Analyzing perceptual attributes of 3D video.* IEEE Transactions on Consumer Electronics, vol. 55, no. 2, pages 864–872, 2009. (Referenced on pages 53 et 79.)

[Yeh 1990] Y Yeh et L Silverstein. *Limits of fusion and depth judgment in stereoscopic color displays.* Human Factors, vol. 32, no. 1, pages 45–60, 1990. (Referenced on pages 22 et 61.)

[You 2010] J You, L Xing, A Perkis et X Wang. *Perceptual Quality Assessment for Stereoscopic Images Based on 2D Image Quality Metrics and Disparity Analysis.* In Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM - 2010, pages 1–6, 2010. (Referenced on pages 80 et 81.)

[Young 1993] M Young, M Landy et L Maloney. *A perturbation analysis of depth perception from combinations of texture and motion cues.* Vision Research, vol. 33, no. 18, pages 2685–96, December 1993. (Referenced on page 28.)

[Zone 2005] Ray Zone. 3-D Filmmakers: Conversations with Creators of Stereoscopicmotion Pictures. Oxford: The Scarecrow Press, 2005. (Referenced on page 13.)

[Zone 2007] Ray Zone. Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952. University Press of Kentucky, 2007. (Referenced on pages 9, 10, 11 et 12.)

Appendix A

Datasets

For the specific needs of our experimental studies, we have created three different datasets of stereoscopic images, using the Fujifilm FinePix Real 3D W3 stereoscopic camera presented in Section 6.1.

The three datasets are available for download on the web site dedicated to the MOOV3D project[1], hosted by our laboratory's server.

The characteristics of each one of these datasets are described next.

A.1 MOOV3D – GIPSA Stereoscopic Dataset 1

The stereoscopic dataset published online under this name is made up of a selection of 24 stereoscopic images that cover various levels of structural complexity, various real world photographed distances, and various levels of binocular disparities. A preview of this selection of images, along with some of their properties, is shown in Table A.1.

However, these 24 stereoscopic images published are only a part of a larger original dataset that we present here. The criteria of classification and selection of the reduced dataset of Table A.1 from the initial larger collection is discussed in this manuscript in the context of the exploratory study that we performed, in Section 8.2.

The initial test material was thus composed of a larger database of 158 stereoscopic images, all taken with the Fujifilm FinePix Real 3D W3 camera presented in the previous chapter at its best quality configuration and at the resolution of 1080p. The minimal focal distance of 6.3 cm was imposed for every image capture, aiming for uniformity in the geometrical conditions of the shooting, given the fact that the stereoscopic camera allowed varying the focal distance.

The images in the dataset were meant to represent situations that ordinary people (non-specialist) would photograph if they owned a compact stereoscopic camera or a mobile phone with a stereoscopic camera. Thus, their content was represented by everyday life sequences like objects, interiors, touristic sites, landscapes, people, or animals. All the images were intended to be as natural (*i.e.* as untreated) as possible, placing the dataset in a consumer utilization context.

For structuring the database, three image characteristics were considered as the three independent variables to define our image collection:

[1] http://www.gipsa-lab.fr/projet/MOOV3D/

Table A.1: A preview of the images in the MOOV3D – GIPSA Stereoscopic Dataset 1.

Image 1	Image 2	Image 3	Image 4	Image 5	Image 6
					
High complexity Average depth interval	High complexity Small depth interval	High complexity –	High complexity Average depth interval	High complexity Small depth interval	High complexity Average depth interval
Image 7	Image 8	Image 9	Image 10	Image 11	Image 12
					
High complexity Average depth interval	Average complexity Large depth interval	High complexity Large depth interval	Low complexity Large depth interval	Average complexity Average depth interval	High complexity Small depth interval
Image 13	Image 14	Image 15	Image 16	Image 17	Image 18
					
High complexity Small depth interval	Average complexity Average depth interval	Average complexity Large depth interval	Average complexity Small depth interval	Average complexity Small depth interval	Low complexity Small depth interval
Image 19	Image 20	Image 21	Image 22	Image 23	Image 24
					
Low complexity Average depth interval	Average complexity Small depth interval	Low complexity Small depth interval	Average complexity Average depth interval	High complexity Large depth interval	High complexity Large depth interval

(1) *complexity* – given by the number of objects in the image, by the textures, the shapes, the contrasts, the colors;

(2) *depth interval* – given by the distance between the closest and the furthest object in the image;

(3) *monocular depth cues* – described later on the experiment task sheet as: "elements in a 2D image that help the observer mentally reconstitute the image in three-dimensions after watching it only in 2D (elements that offer information on the display of the objects in space, on the distances, on the dimensions, on the depths, *etc.*)".

For each characteristic considered, several conditions were possible:

(1) for *complexity*:

- low complexity,
- average complexity,
- high complexity,

(2) for *depth interval*:

- small depth interval,
- average depth interval,
- large depth interval,

(3) for *monocular depth cues*:

- monocular depth cues present,
- monocular depth cues absent,

and, at capture, a lot of attention was payed to covering as well as possible every combination of conditions.

The intersection of the criteria enumerated led to a total of 18 complex categories of images, each such category being defined by one condition of each variable.

Given these considerations, the organization of the database can be represented with a structured three-dimensional matrix, as in Figure A.1. Each of the 18 cases of the matrix acts as a container for the images that satisfy the criteria defined by the position of that case, relative to the coordinate system defined by:

the X axis: the three conditions of the *complexity* variable;

the Y axis: the three conditions of the *depth interval* variable;

the Z axis: the two conditions of the *monocular depth cues* variable.

However, for an easier visualization of the database structure during the experiment and during the data processing phase, a simplified two-dimensional representation of the structure of the database was used. The Z dimension was compressed and each case situated at the intersections of the X and Y axis was separated in two, as shown by Figure A.2.

In this manner, the MOOV3D – GIPSA Stereoscopic Dataset 1 made of 158 stereoscopic images was the stimuli set used in our exploratory study, presented in Chapter 8.

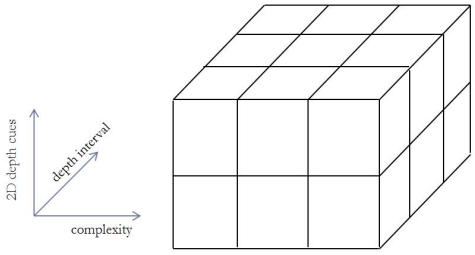


Figure A.1: The schema of the structure of database 1.

3D images database schema	Low complexity	Average complexity	High complexity
Small depth interval	Monocular cues present 1 / 2 Monocular cues absent	Monocular cues present 3 / 4 Monocular cues absent	Monocular cues present 5 / 6 Monocular cues absent
Average depth interval	Monocular cues present 7 / 8 Monocular cues absent	Monocular cues present 9 / 10 Monocular cues absent	Monocular cues present 11 / 12 Monocular cues absent
Large depth interval	Monocular cues present 13 / 14 Monocular cues absent	Monocular cues present 15 / 16 Monocular cues absent	Monocular cues present 17 / 18 Monocular cues absent

Figure A.2: The simplified schema of the structure of database 1.

A.2 MOOV3D – GIPSA Stereoscopic Dataset 2

At the basis of the second stereoscopic dataset that we created, there is a group of 4 still images of different test scenes: two outdoor and two indoor. They are shown in Figure A.2. As it can be noticed, in each indoor or outdoor group, one scene contains only still objects and the other scene a human character, in order to cover a mixed semantic space, despite the reduced number of photographed contexts. Also, the scenes were intentionally chosen of varying spatial information complexity.

The dataset images were taken at the resolution of 1080p with the same compact Fujifilm FinePix Real 3D W3 camera described in Section 6.1, at its finest quality configuration and also by maintaining the 6.3 cm focal distance for all the images. Like for the previous dataset, all the images were intended to be as natural as possible, as taken by ordinary users, placing these stereoscopic images in a consumer utilization context.

Table A.2: The 4 scenes at the basis of the MOOV3D – GIPSA Stereoscopic Dataset 2.

Scene 1	Scene 2	Scene 3	Scene 4
			

The creation of this stimuli set was in accordance with the ITU recommendations [ITU 2012b]. Consequently, for each of the four scenes, we derived 27 test conditions, leading to a set of 108 distinct presentations. We preferred the configuration explained in Equation 4.2 for the definition of the test conditions. Accordingly, the $27 = 3 \times 3 \times 3$ test conditions were represented by the combination of three levels of variations for three independent variables, as follows:

(1) 3 different levels of *circular averaging blur*,

(2) 3 different levels of *the largest crossed binocular disparities*,

(3) 3 different levels of *camera-scene distance at capture*, conveniently chosen in function of each test scene.

The exact values used for the design of the test conditions are detailed in Table A.3. The justification of the choice of these parameters will be however given later, in Section 9.1.1, where the exact context of utilization of this dataset will be presented.

Table A.3: The test conditions considered for building the MOOV3D – GIPSA Stereoscopic Dataset 2 (the values given for the blur indicate the *radius*, which allows defining the size in pixels of the *pillbox* filter to be applied, size computed as a square of side $2 \times radius + 1$; the disparities are expressed in minutes of arc).

blur	disparities	capture distance			
		scene 1	scene 2	scene 3	scene 4
0	[-21,-4]	6.26 m	10.25 m	28.5 m	5.5 m
1	[-86,-53]	7.76 m	15 m	31.6 m	10 m
1.75	[-140,-104]	9.26 m	20 m	36 m	15 m

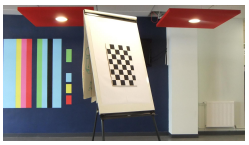
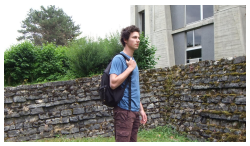
Broadly speaking, the MOOV3D – GIPSA Stereoscopic Dataset 2 was designed having in view the quantitative experiment described in Section 9.1 and the design of the test conditions was done with the purpose of refining the findings of our exploratory study into a stereoscopic image quality model.

A.3 MOOV3D – GIPSA Stereoscopic Dataset 3

The third dataset built for our experiments was created using the same principles as the MOOV3D – GIPSA Stereoscopic Dataset 2. All the points illustrated in Section A.2 are valid for this database as well, with two exceptions: the base scenes and the test conditions.

The foundation of this third dataset is represented by the 8 scenes shown in Figure A.4. Like for the dataset presented previously, they also represent exterior and interior scenes in equal proportions. As well, they contain both scenes containing only objects and scenes with human characters.

Table A.4: The 8 scenes at the basis of the MOOV3D – GIPSA Stereoscopic Dataset 3.

Scene 1	Scene 2	Scene 3	Scene 4
			
Scene 5	**Scene 6**	**Scene 7**	**Scene 8**
			

For each of the 8 scenes in Figure A.4, a number of 9 test conditions were derived, generating a final dataset of 72 distinct presentations. The configuration in Equation 4.2 was again the one preferred for the design of the stimuli set and the $9 = 3 \times 3$ test conditions were represented this time by the combination of three levels of variations for only two independent variables, as follows:

(1) 3 different values of *focal distance*,

(2) 3 different *main-subject-to-background distances*.

Details on the test conditions considered are in Table A.5. Like for the previous stimuli set, the explanation of the reasons that led to this dataset structure will be given later, in the context of its utilization, in Section 9.1.1.

Limiting the number of test conditions to only 9, compared to 27 previously, allowed us to augment the number of base scenes from 4 to 8. This illustrates how the specificity of one experiment imposes constraints on the dataset to be used.

The choice of the configuration illustrated in Equation 4.2 for both Datasets 2 and 3 was justified by the fact that in the case of all our quantitative experiments, we were not only interested by the individual effect of the considered independent variables on the subjective ratings, but also by their joint effect.

Table A.5: The test conditions considered for building the MOOV3D – GIPSA Stereoscopic Dataset 3.

focal distance	object-background distance
6.3 mm	4 m
13.1 mm	8 m
18.9 mm	12 m

The MOOV3D – GIPSA Stereoscopic Dataset 3 has an additional particularity. For the 18.9 mm focal distance, the consumer Fujifilm stereoscopic camera produces a slight blur, despite using a tripod and the camera timer. Therefore, since we intended to avoid as much as possible the influence of other uncontrolled variables on the structure of our database, we evaluated the level of blur in all the images of the original database, observed the elevated blur for the largest focal distance condition, and then we added blur on the sharper images, such as the level of blur across the database would be as uniform as possible. The objective measures were effectuated using the no-reference blur metric proposed in [Crete 2007], which is described in Appendix E.1.

Consequently, two versions of the dataset are available online: one version with the stereoscopic images in the form in which the stereoscopic camera rendered them and one version in which the slight blur was added to the sharp images.

This third dataset was also designed for the purpose of a quantitative experiment. Its utilization is described in Section 9.1.

Appendix B

Statistical Tools

B.1 Processing the experimental data

Main reference: [ITU 2012b].

During an experimental quantitative study following the single stimulus (SS) method described in Section 4.7.3.4, a series of *subjective ratings* are recorded. These are the ratings that all the participants give to each of the stimuli of the dataset under study. Thus, a matrix of $N \times J \times K \times R$ ratings is collected per experiment, where N is the number of participants to that experiment, J is the number of test conditions applied to the K basic images of the dataset (leading to a dataset of $J \times K$ stimuli), and R is the number of times each stimulus was displayed for rating.

The ratings can take various types of values. This Appendix, however, proposes solutions adequate to the particular type of experimental methods which collect *numeric ratings*. In the case of our experiments, these were integer values between 1 and 5.

In every rating collection obtained after an experiment, there are *variations in the distributions of the votes*. These are due to the differences in judgment between observers and to the effect of a variety of conditions associated with the experiment.

Therefore, statistical methods are used for concentrating all the ratings in a *unique value* per stimulus, corresponding to the general tendency of voting for that specific stimulus.

We present here such a method, in the form of the computation of mean opinion scores and of their corresponding confidence intervals.

B.1.1 Mean opinion scores (MOS)

A mean opinion score can be computed for each of the stimuli in the dataset used.

Thus, for an experimental stimuli corresponding to the test condition j of image k, displayed during the repetition r, the corresponding *mean opinion score* is computed as:

$$\overline{u_{jkr}} = \frac{1}{N} \sum_{i=1}^{N} u_{ijkr}, \tag{B.1}$$

where u_{ijkr} is the score of the observer i for the test condition j of the image k, during repetition r.

Similarly, overall mean scores, $\overline{u_j}$ or $\overline{u_k}$, can be computed for each test condition or for each distinct image of the dataset.

B.1.2 Confidence intervals

In order to accompany the computed mean opinion scores with a measure of reliability of their estimation, the *confidence intervals* are proposed. These are derived from the standard deviation and size of each sample.

A 95 % *confidence interval* is of the form:

$$[\overline{u_{jkr}} - \delta_{jkr}, \overline{u_{jkr}} + \delta_{jkr}], \tag{B.2}$$

where:

$$\delta_{jkr} = 1.96 \cdot \frac{S_{jkr}}{\sqrt{N}} \tag{B.3}$$

and S_{jkr} is the standard deviation for each stimulus, computed as:

$$S_{jkr} = \sqrt{\sum_{i=1}^{N} \frac{(\overline{u_{jkr}} - u_{ijkr})^2}{N-1}}. \tag{B.4}$$

The significance of a 95 % confidence interval associated to a computed mean score is that, with a probability of 95 %, the absolute value of the difference between the computed experimental mean score and the "true" mean score (for a very high number of observers) is smaller than this interval.

B.2 The β_2 test (computing the kurtosis coefficient)

Main reference: [ITU 2012b].

The β_2 test can be used in order to ascertain whether a *distribution of scores* can be considered *normal*. This verification can be performed by computing the *kurtosis coefficient* of that distribution, *i.e.* the ratio of the fourth order moment to the square of the second order moment, as:

$$\beta_{2_{jkr}} = \frac{m_4}{(m_2)^2}, \tag{B.5}$$

where:

$$m_x = \frac{\sum_{i=1}^{N}(u_{ijkr} - \overline{u_{ijkr}})^x}{N}. \tag{B.6}$$

If the β_2 coefficient computed is between 2 and 4, the distribution may be taken to be normal.

B.3 Screening of the observers

Main reference: [ITU 2012b].

The *ITU screening method* for the single stimulus (SS) experiments has the purpose of discarding from the experimental data the scores corresponding to observers who voted differently from the general tendency of the totality of the observers.

This procedure should not be applied more than once to the results of a given experiment. Moreover, the use of the procedure should be restricted to cases in which there are relatively few observers (*e.g.* fewer than 20), all of whom are non-experts.

In order to select the observers to be discarded, the next procedure must be followed.

For each presentation, the scores u_{ijkr} of each observer must be compared with their associated mean value, $\overline{u_{jkr}}$ plus their associated standard deviation S_{jkr}, times 2 if the distribution of the scores is normal, or times $\sqrt{20}$ if the distribution of the scores is not normal. This comparison term can be denoted by P_{jkr}. The same scores of each observer must also be compared with their associated mean value minus their associated standard deviation times 2 or times $\sqrt{20}$, term that is denoted by Q_{jkr}.

Every time a score of an observer is found above P_{jkr}, a counter associated with that particular observer, P_i, must be incremented. Similarly, every time a score of an observer is found below Q_{jkr}, a counter associated with that observer, Q_i, must be incremented.

These steps can be mathematically expressed as:

For each observer i, find P_i and Q_i, such as, for $j, k, r = 1, 1, 1$ *to* J, K, R:

$$\begin{aligned}
&if\ 2 \leq \beta_{2_{jkr}} \leq 4,\ then: \\
&\qquad if\ u_{ijkr} \geq \overline{u_{jkr}} + 2 \cdot S_{jkr}, && then\ P_i = P_i + 1, && \text{(B.7a)} \\
&\qquad if\ u_{ijkr} \leq \overline{u_{jkr}} - 2 \cdot S_{jkr}, && then\ Q_i = Q_i + 1, && \text{(B.7b)} \\
&\quad else: \\
&\qquad if\ u_{ijkr} \geq \overline{u_{jkr}} + \sqrt{20} \cdot S_{jkr}, && then\ P_i = P_i + 1, && \text{(B.7c)} \\
&\qquad if\ u_{ijkr} \leq \overline{u_{jkr}} - \sqrt{20} \cdot S_{jkr}, && then\ Q_i = Q_i + 1, && \text{(B.7d)}
\end{aligned}$$

where:

$$P_{jkr} = \overline{u_{jkr}} + 2 \cdot S_{jkr} \tag{B.8a}$$

or

$$P_{jkr} = \overline{u_{jkr}} + \sqrt{20} \cdot S_{jkr} \tag{B.8b}$$

and

$$Q_{jkr} = \overline{u_{jkr}} - 2 \cdot S_{jkr} \tag{B.9a}$$

or

$$Q_{jkr} = \overline{u_{jkr}} - \sqrt{20} \cdot S_{jkr}. \tag{B.9b}$$

Finally, the following two ratios must be calculated: $P_i + Q_i$ divided by the total number of scores from each observer for the whole session and $P_i - Q_i$ divided by $P_i + Q_i$ as an absolute value. If the first ratio is greater than 5 % and the second ratio is less than 30 %, then observer i must be eliminated, as also shown mathematically:

$$if \qquad \frac{P_i + Q_i}{J \cdot K \cdot R} > 0.05 \qquad and \qquad |\frac{P_i - Q_i}{P_i + Q_i}| < 0.3, \tag{B.10}$$

then *reject observer i*.

B.4 Kendall's W

Main reference: [Chanquoy 2005].

Kendall's W, also called *Kendall's coefficient of concordance*, is a non-parametric statistic that can be used for testing the level of agreement among the raters of a same dataset.

By definition, Kendall's W is a measure of agreement among the ranks of multiple classifications and can be computed with the following formula:

$$W = \frac{12 \cdot S}{k^2 \cdot (n^3 - n)}, \tag{B.11}$$

where 12 is a constant, k is the number of raters, n is the number of rated stimuli, and S is the sum of the squared differences between any total rank per stimulus and the mean of the total ranks of all stimuli, as shown:

$$S = \sum_{i=1}^{N} (R_i - \overline{R})^2, \tag{B.12}$$

with R_i representing the total of the ranks given to the stimulus i and $\overline{R}$ representing the mean values of all the total ranks.

Kendall's W coefficient can vary between -1 and 1, where -1 corresponds to a maximum negative correlation, and 1 corresponds to a maximum positive correlation between the two sets of data considered.

B.5 Scores normalization

Even if an experimental study is performed with care and the use of the rating scale is correctly explained to the participants, the final results might show important variations in the way the scale was employed by different persons.

In order to compensate for these differences and to adjust the scores to a scale common in meaning for all the participants, a *normalization* step is often performed before any other step in the processing of the experimental data.

As a normalization solution for our data, we preferred the simple solution of transforming the original subjective scores into *standard scores* (or *z-scores*). This was performed with the following formula:

$$z_{ijkr} = \frac{u_{ijkr} - \overline{u_{ir}}}{\sigma_{ir}} \tag{B.13}$$

for every score u_{ijkr} of the participant i on the test condition j of image k during repetition r, where $\overline{u_{ir}}$ represents the mean of all the ratings of participant i during repetition r and σ_{ir} represents the standard deviation of these ratings.

After the normalization step, a *remapping* of the computed z-scores to the original interval of the voting scale may be performed or not. We chose to perform this remapping in order to present the final results relative to the original scale used.

B.6 Analysis of variance (ANOVA)

Main reference: [Chanquoy 2005].

The ANOVA statistical procedure allows the comparison of estimated variances, *i.e.* allows to the experimenter to test whether the scores obtained (*i.e.* the dependent variables) for different groups vary or not significantly. For this, ANOVA does not only take into account the mean values, but also the variances, *i.e.* the distribution of the scores.

The procedure to follow in order to perform an ANOVA statistical test is rigorous and can be schematized in the following three steps:

1. The statistical hypothesis must be set: the null hypothesis and the alternative hypothesis (H_0 and H_1). In the typical application of ANOVA, the null hypothesis states that all the considered groups are random samples of the same population. This implies that all the treatments of the independent variables have the same effect (perhaps none). Rejecting the null hypothesis implies that different treatments result in altered effects.

2. A series of calculations must be performed in order to compare variances, sources of variability; the ratio between two different sources of variability leads to an index F, that we denote here by F_c (the computed F).

3. F_c must be compared to a theoretical F (either given by statistical tables, or computed by the software used). Consequently, if F_c is superior to the theoretical F, or if the probability associated to F_c is sufficiently high, the H_0 hypothesis can be rejected and the independent variables are found to have a significant effect on the dependent variable for the entire population, and not only for the tested sample.

B.7 Wilcoxon test

Main reference: [Chanquoy 2005].

The *Wilcoxon test* is a non-parametric statistical hypothesis test used for comparing the means of two samples in order to assess whether their population mean ranks differ.

The only constraint imposed for applying this test is that the distributions must be measured on an ordinal scale.

The hypotheses set for the Wilcoxon test are similar to those for the ANOVA, however, in this case, the *W test statistic* is the one computed and evaluated.

B.8 Scales aggregation

Main reference: [Marcotorchino 1979].

When, after an experimental study, a table T is obtained, storing the ratings or the ranks attributed by a series of N judges to a set of M stimuli, the result of the experiment is often searched in the form of a set of M values representing the concentrated opinion of all the judges on each of the M stimuli. This problem can be reformulated as a *scale aggregation* problem trying to identify a vector of size M equivalent to the ratings or ranks given by a *virtual judge* to the M stimuli, vector that would *aggregate* the totality of the opinions of the N judges. This would thus be a vector X of positive values that would minimize the following function:

$$min \sum_{i=1}^{N} \mu(X, T_i), \tag{B.14}$$

where μ is a metric that characterizes the proximity between the X rating scale and the T_i rating scales.

The μ metric is set a priori in function of the specificity of the data under study. A series of models are proposed in the literature for this metric. Among them, for the data processing phase of our exploratory test, we preferred to consider the *absolute distance d* metric of the form:

$$d(T_j, T_k) = \sum_{i=1}^{M} |t_{ij} - t_{ik}|, \tag{B.15}$$

which is equivalent to determining the i component of the X vector as equal to the *median* value of all the x_{ik} values, where $k = \overline{1, N}$.

Appendix C

The Simulator Sickness Questionnaire

The form in which the *Simulator Sickness Questionnaire (SSQ)* was originally proposed in [Kennedy 1993] is reproduced here.

Table C.1: Our adaptation of the original SSQ.

	Weight		
SSQ Symptom	N	O	D
General discomfort	1	1	0
Fatigue	0	1	0
Headache	0	1	0
Eyestrain	0	1	0
Difficulty focusing	0	1	1
Increased salivation	1	0	0
Sweating	1	0	0
Nausea	1	0	1
Difficulty concentrating	1	1	0
Fullness of head	0	0	1
Blurred vision	0	1	1
Dizzy (eyes open)	0	0	1
Dizzy (eyes closed)	0	0	1
Vertigo	0	0	1
Stomach awareness	1	0	0
Burping	1	0	0

The method used for computing the N (nausea), O (oculomotor), and D (disorientation) scores, as well as the TS (total score) is the following:

$$N = [1] \times 9.54, \tag{C.1a}$$
$$O = [2] \times 7.58, \tag{C.1b}$$
$$D = [3] \times 13.92, \tag{C.1c}$$
$$TS = [1] + [2] + [3] \times 3.74, \tag{C.1d}$$

where [1], [2], and [3] stand for the sums obtained by adding the weighted scores corresponding to each of the three columns in Table C.1, *i.e.* the N, O, and D columns, respectively.

Appendix D

Experiment Details

D.1 Experiment 1

D.1.1 Experiment script

1. Fill in the participant data form.
2. Have the participant read the task sheet.
3. Part I - Classification.
4. At the end of the classification, fill in the classification sheet.
5. Perform the vision tests:
   - Visual acuity test – Snellen chart,
   - Color vision test – Ishihara test,
   - Stereo acuity test – McGill test.
6. Decide whether the participant can continue or not.
7. Tell the participant to choose a correct and comfortable position.
8. Turn off the main lights.
9. Check the correct functioning of the display system with a test image.
10. Launch the experiment software.
11. Start recording.
12. Part II - Visualization and verbalization.
13. Make a back-up of the recording file.

D.1.2 Task sheet

1. Classify 2D images:
   - watch the printed 2D images and arrange them on the table in front of you, in function of the three classification categories considered:
     - complexity - given by the number of objects in the image, by the textures, the shapes, the contrasts, the colors,

  - depth interval - given by the distance between the closest and the furthest object in the image,
  - 2D depth cues - elements present in a 2D image that help the observer to mentally reconstitute the image in three-dimensions only after watching it in 2D (elements that offer information on the display of the objects in space, on the distances, on the dimensions, on the depths, etc.); examples of 2D depth cues:
    * lighting and shades,
    * occlusion,
    * relative size,
    * linear perspective,
    * texture gradient,
    * relative density (atmospheric blur, saturation, color intensity variation),
    * depth of field,
    * position relative to the horizon,
    * presence of known size objects.

2. Watch and evaluate 3D images:

   · watch each 3D image displayed on the TV screen and freely express your opinion on its quality and on the perceived comfort/discomfort;

   · freely choose the visualization time for each picture.

D.2 Experiments 2 and 3

D.2.1 Experiment script

1. Turn the day light lamp on.
2. Check the HDMI connections.
3. Read the *Information notice* with the participant if he/she has not done it already.
4. Have the *Consent form* signed by the participant.
5. Hand to the participant a copy of the *Information notice* and a copy of the *Consent form* with both signatures on.
6. Fill in the form on the participant data.
7. Perform the vision tests:

   · Interocular distance (measure),

   · Visual acuity test – Snellen chart,

   · Color vision test – Ishihara test,

   · Stereo acuity test – TNO test.

8. Decide whether the participant can continue or not.

9. Have the Simulator Sickness Questionnaire filled in by the participant (*only for Experiment 2*).
10. Have the participant read the corresponding task sheet.
11. Tell the participant to choose a correct and comfortable position.
12. Turn off the main lights.
13. Let the participant adjust the 3D glasses to his/her view using the calibration image (*only for the sub-experiments on the 3D glasses.*).
14. Check the correct functioning of the display system with a test image.
15. Launch the experiment software.
16. Make a back-up of the results file.
17. Have the Simulator Sickness Questionnaire filled in again by the participant (*only for Experiment 2*).

D.2.2 Task sheets

D.2.2.1 Sub-experiments on image quality

This experiment will include three parts:

1. The learning phase,
2. The evaluation of the first group of images,
3. The evaluation of the second group of images.

During each part of the experiment, you will be shown sequences of three different elements:

· a white cross on a grey background – you must concentrate your gaze on it!

· a 3D image – you must explore it attentively!

· the text "Vote now!" on a grey background – you must give a rating on the *image quality* of the 3D scene that was previously displayed!

By *image quality* we mean the 2D quality of the 3D image presented, *i.e.* the level of quality that the 3D image would have if it was presented in 2D instead of 3D.

The rating on each image should take a value between 1 and 5, where:

5 image quality: excellent

4 image quality: good

3 image quality: fair

2 image quality: poor

1 image quality: bad

The rating for a 3D image can be given only while the "Vote now!" text is being displayed and it should be recorded using the highlighted keys on the keyboard. As soon as a rating is given, a new image will be automatically displayed (always preceded by a white cross).

D.2.2.2 Sub-experiments on comfort

This experiment will include three parts:

1. The learning phase,
2. The evaluation of the first group of images,
3. The evaluation of the second group of images.

During each part of the experiment, you will be shown sequences of three different elements:

- a white cross on a grey background – you must concentrate your gaze on it!
- a 3D image – you must explore it attentively!
- the text "Vote now!" on a grey background – you must give a rating on the *image quality* of the 3D scene that was previously displayed!

By *comfort* we mean the general state of comfort induced by the 3D image presented.

The rating on each image should take a value between 1 and 5, where:

5 comfort: excellent

4 comfort: good

3 comfort: fair

2 comfort: poor

1 comfort: bad

The rating for a 3D image can be given only while the "Vote now!" text is being displayed and it should be recorded using the highlighted keys on the keyboard. As soon as a rating is given, a new image will be automatically displayed (always preceded by a white cross).

D.2.2.3 Sub-experiments on realism

This experiment will include three parts:

1. The learning phase,
2. The evaluation of the first group of images,
3. The evaluation of the second group of images.

During each part of the experiment, you will be shown sequences of three different elements:

- a white cross on a grey background – you must concentrate your gaze on it!
- a 3D image – you must explore it attentively!
- the text "Vote now!" on a grey background – you must give a rating on the *image quality* of the 3D scene that was previously displayed!

By level of *realism* we mean the level of accuracy of the 3D rendering, *i.e.* the level of resemblance between the 3D image displayed and what you would see in reality if you were in the spot from where the picture was taken and you saw the real scene.

The rating on each image should take a value between 1 and 5, where:

5 realism: excellent

4 realism: good

3 realism: fair

2 realism: poor

1 realism: bad

The rating for a 3D image can be given only while the "Vote now!" text is being displayed and it should be recorded using the highlighted keys on the keyboard. As soon as a rating is given, a new image will be automatically displayed (always preceded by a white cross).

D.2.2.4 Sub-experiments on overall 3D quality

This experiment will include three parts:

1. The learning phase,
2. The evaluation of the first group of images,
3. The evaluation of the second group of images.

During each part of the experiment, you will be shown sequences of three different elements:

- a white cross on a grey background – you must concentrate your gaze on it!
- a 3D image – you must explore it attentively!
- the text "Vote now!" on a grey background – you must give a rating on the *image quality* of the 3D scene that was previously displayed!

By *overall 3D quality* we mean the percept that includes the quality of all the aspects related to the 3D image. This percept should not exclusively include the quality that the image would have if it was displayed in 2D, but also the quality rendered by the fact that the image is presented in three dimensions.

The rating on each image should take a value between 1 and 5, where:

5 overall 3D quality: excellent

4 overall 3D quality: good

3 overall 3D quality: fair

2 overall 3D quality: poor

1 overall 3D quality: bad

The rating for a 3D image can be given only while the "Vote now!" text is being displayed and it should be recorded using the highlighted keys on the keyboard. As soon as a rating is given, a new image will be automatically displayed (always preceded by a white cross).

Appendix E

Objective measures

E.1 Measuring the blur

In order to compute an objective score representing the level of blur evaluated for a given stereoscopic image, we used the *no-reference perceptual blur metric* presented in [Crete 2007].

This metric estimates the level of the *blur annoyance* of a 2D image based on the discrimination between different levels of blur that are perceptible on the same picture. It uses the fact that the humans have difficulties in perceiving differences between an originally blurred image and the same image blurred once more.

Therefore, the algorithm blurs the image under study and makes a decision on the intensity of its blur level in function of the variations between the original image and its blurred copy. If the variations are large, the original image can be considered of good quality (no or little amounts blur). If the variations are small, it can be implied that the original image already contained a considerable level of blur.

The no-reference perceptual blur metric only takes into account the pixels that changed after the blurring step, in order to cope with the situations in which only a minor part of the image is blurred. Also, it only considers the neighboring pixels variations which have decreased after the blurring step, in order to avoid the zones where new slight pixels variations that did not exist before appear after blurring for example a sharp edge on a flat area.

The blur is estimated with this metric both in the horizontal and the vertical directions and the more annoying blur from the two is considered as the final blur value.

The flow-chart describing the functioning of this metric is reproduced in Figure E.1.

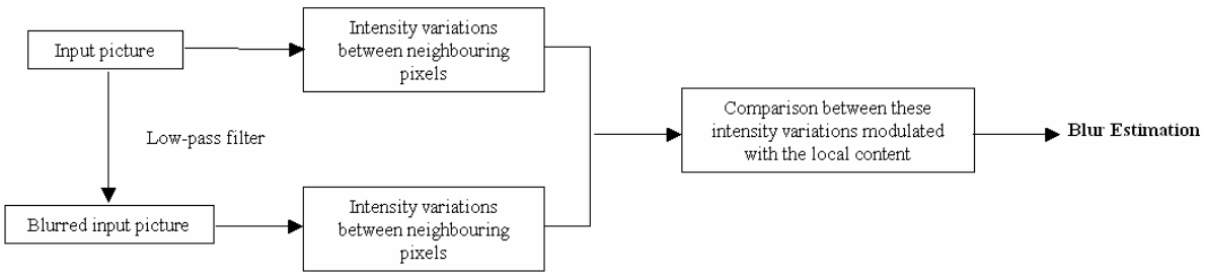


Figure E.1: The simplified flow-chart of the no-reference perceptual blur metric principle [Crete 2007].

The values produced by the no-reference blur metric range between 0 and 1, with 0 representing the best quality and 1 representing the worst quality in terms of blur perception.

For our stereoscopic images, we applied the blur metric separately on each of the two views of a stereoscopic pair and the *mean* of the two values obtained was considered the blur measure of the stereoscopic image.

E.2 Measuring the disparities

In order to determine the *binocular disparities* contained by a given stereoscopic image, we used a Matlab stereo-matching algorithm, based on a C++ implementation [Lowe 2013] of the *SIFT feature detection method* described in [Lowe 1999, Lowe 2004].

This algorithm selects a large number of SIFT feature points in the stereo pairs, matches them, then returns the coordinates of the points in each of the views of the stereo pairs. Figure E.2 gives an example of such a stereo-match obtained.

Figure E.2: Example of a stereo-match determined with our algorithm based on the SIFT feature detection method.

The totality of these coordinates allowed us to compute all the *horizontal and vertical disparities* detected, then to make a series of meaningful representations or calculations. Consequently, horizontal and vertical disparities histograms could be created, as exemplified in Figure E.3, or the largest crossed or uncrossed disparities could be computed for any stereoscopic image.

Vertical disparities could be noticed in the majority of the disparity histograms of the stereoscopic images taken with our stereoscopic camera and their presence was justified by the small optical imperfections of the camera lenses, which lead to slight vertical displacements of the corresponding points in the two views. The magnitude of these vertical disparities was however each time far under the threshold of uncomfortable vertical fusion of the human visual system, therefore their presence was of no inconvenience in our study and no geometrical corrections were applied to our datasets.

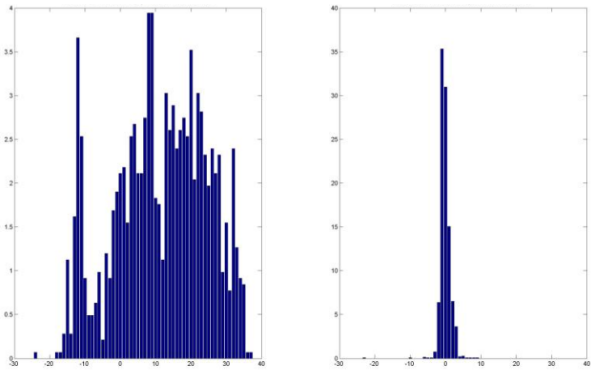


Figure E.3: Examples of a horizontal disparities and a vertical disparities histogram for a stereoscopic image (examples corresponding to the image in Figure E.2.

Multiple representations of the calculated disparities were possible. Among them, a meaningful representation that we preferred in our interpretations was that of the horizontal disparities in function of the x coordinate of the matched key points for which they were calculated. This kind of representation could be compared to seeing a three-dimensional representation of the key points in the real scene from above and an example is given in Figure E.4.

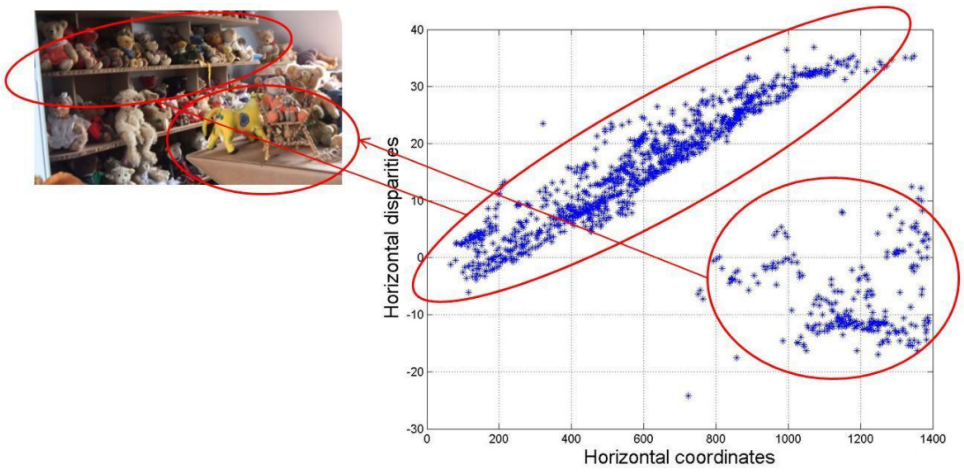


Figure E.4: Example of representation of the horizontal disparities in a stereoscopic image in function of the x coordinates of the matched key points for which they were computed (example corresponding to the image in Figure E.2.

All the computations in our work related to the disparities of a stereoscopic image were performed using the method described.

E.3 Measuring the cardboard effect

Given three points $P_{0_1}(x_{0_1}, y_{0_1}, z_{0_1})$, $P_{0_2}(x_{0_2}, y_{0_2}, z_{0_2})$, and $P_{0_3}(x_{0_3}, y_{0_3}, z_{0_3})$ in the real space, with $P_{0_1}P_{0_2}$ parallel to the line determined by the positions of the lenses of a stereoscopic camera and P_{0_3} at a different depth than $P_{0_1}P_{0_2}$ (therefore $z_{0_1} = z_{0_2}$ and $z_{0_1} \neq z_{0_3}$) and given the three

image points $P_{i_1}(x_{i_1}, y_{i_1}, z_{i_1})$, $P_{i_2}(x_{i_2}, y_{i_2}, z_{i_2})$, and $P_{i_3}(x_{i_3}, y_{i_3}, z_{i_3})$ that a person with an eye separation e, placed at a distance D from the screen, would perceive if the parameters of the stereoscopic system are:

M: the scaling factor from the sensor to the screen;

f: the stereoscopic camera focal distance;

t: the stereoscopic camera lens separation,

then the *cardboard effect ratio* can be computed as:

$$cdb\_eff\_ratio = \frac{represented\_width/represented\_depth}{real\_width/real\_depth} = \frac{(x_{i_2} - x_{i_1})/(z_{i_3} - z_{i_1})}{(x_{0_2} - x_{0_1})/(z_{0_3} - z_{0_1})}. \quad \text{(E.1)}$$

The values obtained for this ratio can be interpreted by comparison to the ideal ratio of 1 which means that the proportions in width and depth are maintained between the real world scene and the perceived image space. If the cardboard effect ratio is smaller or larger that 1, this means that the represented depths are compressed or dilated relative to the original proportions in the real scene.

Also, given the mathematical description of the geometry of our stereoscopic system in Section 6.4, the cardboard effect ratio can also be written as:

$$cdb\_eff\_ratio = \frac{e \cdot z_{0_3} + M \cdot f \cdot t + M \cdot z_{0_3} \cdot (p_{xl} - p_{xr})}{D \cdot t}. \quad \text{(E.2)}$$

Using this formula, *a theoretical value of the cardboard effect ratio* corresponding to a certain object in a stereoscopic image can be computed, if the distance between that object and the stereoscopic camera at capture is known.

Résumé — Dans le contexte d'un intérêt grandissant pour les systèmes stéréoscopiques, mais sans méthodes reproductibles pour estimer leur qualité, notre travail propose une contribution à la meilleure compréhension des mécanismes de perception et de jugement humains relatifs au concept multi-dimensionnel de *qualité d'image stéréoscopique*. Dans cette optique, notre démarche s'est basée sur un certain nombre d'outils : nous avons proposé un cadre adapté afin de structurer le processus d'analyse de la qualité des images stéréoscopiques, nous avons implémenté dans notre laboratoire un système expérimental afin de conduire plusieurs tests, nous avons créé trois bases de données d'images stéréoscopiques contenant des configurations précises et enfin nous avons conduit plusieurs expériences basées sur ces collections d'images. La grande quantité d'information obtenue par l'intermédiaire de ces expérimentations a été utilisée afin de construire un premier modèle mathématique permettant d'expliquer la perception globale de la qualité de la stéréoscopie en fonction des paramètres physiques des images étudiées.

Mots clés : qualité des images stéréoscopiques, qualité d'image, confort, réalisme, expériences subjectives, expériences psycho-perceptuelles, expériences qualitatives, études exploratoires, expériences quantitatives, metriques objectives, facteurs humains, affichage stéréoscopique, TV 3D, lunettes 3D.

Abstract — In a context of ever-growing interest in stereoscopic systems, but where no standardized algorithmic methods of stereoscopic quality assessment exist, our work stands as a step forward in the understanding of the human perception and judgment mechanisms related to the multidimensional concept of *stereoscopic image quality*. We used a series of tools in order to perform in-depth investigations in this direction: we proposed an adapted framework to structure the process of stereoscopic quality assessment, we implemented a stereoscopic system in our laboratory for performing various tests, we created three stereoscopic datasets with precise structures, and we performed several experimental studies using these datasets. The numerous experimental data obtained were used in order to propose a first mathematical framework for explaining the overall percept of stereoscopic quality in function of the physical parameters of the stereoscopic images under study.

Keywords: stereoscopic image quality, image quality, comfort, realism, subjective experiments, psycho-perceptual experiments, qualitative experiments, exploratory studies, quantitative experiments, objective metrics, human factors, stereoscopic display, 3D TV, 3D glasses.

GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus BP 46,
F-38402 Saint Martin d'Hères CEDEX